UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

FACULTAT D'INFORMÀTICA DE BARCELONA

MASTER IN INNOVATION AND RESEARCH IN INFORMATICS

DATA MINING AND BUSINESS INTELLIGENCE

MASTER THESIS

Predicting Financial Distress Through Machine Learning

Author:
Malika IBRAIMOVA

Supervisor:
Alfredo VELLIDO

A thesis submitted in fulfillment of the requirements for the degree of Master in Innovation and Research in Informatics

Computer Science Department

January 19, 2019

Declaration of Authorship

I, Malika IBRAIMOVA, declare that this thesis titled, “Predicting Financial Distress Through Machine Learning” and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

Abstract

The beforehand identification of future financial distress of a company might be very helpful for managers, stockholders, creditors and other interested third parties to discover the financial health of company more deeply. The main question which will be raised in this thesis is - whether we can predict future financial distress of a company based on the changes in historical financial results using different machine learning techniques. The predictions were made based on changes in financial results during three different time intervals, which are: one year, half-year and a quarter before expected bankruptcy. The financial data of banks used in analysis was obtained from the quarterly reports presented on the website of the Federal Deposit Insurance Corporation. The results of analysis indicated that classification model developed by RBF kernel SVM using the data, obtained from PCA analysis on the basis of quarterly changes of the financial data, best predicts future financial distress in banks of Unites States of America.

Acknowledgements

I would like to thank my supervisor, Alfredo Vellido, for his assistance and guidance which facilitated my work. A special thanks to my family for providing me with endless support throughout my university studies. . . .

Contents

List of Figures

List of Tables

List of Abbreviations

CFA	**C**onfirmatory **F**actor **A**nalysis
EFA	**E**xploratory **F**actor **A**nalysis
FA	**F**actor **A**nalysis
FDIC	**F**ederal **D**eposit **I**nsurance **C**orporation
ICA	**I**ndependent **C**omponent **A**nalysis
KMO	**K**aiser **M**eyer **O**lkin
KNN	**K**- **N**earest **N**eighbors
ML	**M**aximum **L**ikelihood
PC	**P**rincipal **C**omponent
PCA	**P**rincipal **C**omponent **A**nalysis
RF	**R**andom **F**orests
SOM	**S**elf - **O**rganizing **M**aps
SVM	**S**upport **V**ector **M**achines
U.S.	**U**nited **S**tates of America
VIF	**V**ariance **I**nflation **F**actors
VSS	**V**ery **S**imple **S**tructure

Chapter 1

Introduction

1.1 Motivation

Bankruptcy is a process that a legal entity or an individual undertakes in order to protect or liquidate assets to repay debts, depending on the situation. Judging by the financial crisis history and multiple valuations of the several banks' behaviors - bankruptcy is not a problem as such, but a measure taken to solve a bigger problem. Many myths surrounding bankruptcy often create a barrier in the human mind to adequately assess the financial situation of the legal entity. When foreseen beforehand, the bankruptcy can actually become a great tool to avoid further damage and total annihilation of personal assets. That is why it is extremely important to be able to predict any possible financial distress in a company and have up-to-date and constantly improving algorithms that will help to analyze the future financial health of an organization through the financial dynamics of the organization.

Financial distress of the legal entity can be caused by numerous factors that are extremely important to be observed. In today's world, every legal entity has to consider the preemptive measures that will help either to foresee the financial distress and try to avoid it, or file for bankruptcy at the right time. Let us bring about the case of the Lehman Brothers investment bank going bankrupt in September of 2008, an institution that was established in 1850. The financial crisis itself did not begin because of Lehman Brothers having to go bankrupt. The financial crisis had been lurking for the more than a year before this bankruptcy came to happen. A large US investment bank, Bear Stearns, had to be saved in March of 2007. The systematic loss of trust amongst financial institutions towards each other's solvency caused the stress. United States Bankers had made huge benefits by selling "risk free" assets, called mortgage-backed securities, which were basically subprime mortgages mixed together with a better-quality mortgage. In 2006, the Central Bank of The United States had raised the interest which made many households to go default, leading to a decrease in house prices, making it clear that mortgage-backed securities were very risky. Risky securities were owned by almost every bank and it was impossible to recognize which bank was more strongly damaged by the bad debts. Because of all the distrust among each other, banks started to charge high rates to other banks that were suspicious to have unrecognized damage from the mortgage backed securities. And it was precisely in this period of distrust when banks stopped to lend to each other that Lehman Brothers went bust[1].

Why did such a huge investment bank had to file for bankruptcy? The answer is a failure to properly predict the financial distress that was about to hit the U.S. economy and subsequently holding on to the problematic assets and low-rated mortgages that amounted to the figure 30 times bigger than its capital. However, there are

[1] https://www.independent.co.uk/news/business/analysis-and-features/financial-crisis-2008-why-lehman-brothers-what-happened-10-years-anniversary-a8531581.html

also cases of a timely financial distress prediction that helped the banks to avoid unnecessary damages. The BNP Paribas bank, formerly controlled by the government and with current headquarters in Paris, France, is the third largest bank in European Union. The BNP Paribas bank was able to avoid the financial distress of 2007-2008 because of its more conservative business structure and forecasting. More than a half of the BNP Paribas's revenue came from retail banking in European Countries like France, Spain and Italy. Due to their substantial accumulated retail deposits that were $659 billion in 2007, BNP Paribas were able to avoid short-term lending and fund its investment banking operations with existing capital[2].

Predicting financial distress is an activity aimed at identifying and studying possible dangers for the future development of a company. For sustainable development, each company must make predictions about its future financial condition. This is due to the fact that predicting financial distress for a company is of great importance, since it helps to reduce various types of risks that affect the financial condition, increase the efficiency of the company, increase profitability and help to avoid bankruptcy of the company. Financial condition prediction is characterized by uncertainty, which is a problem, since for forecasting it is necessary to take into account many factors that may affect the company in the future. However, after analyzing the most likely factors that may have influence on the financial health of the company, the company's future financial condition can be predicted and the threats that may lead to a decrease in financial viability or even bankruptcy can be foreseen.

1.2 Goals and Research Questions

The research in this thesis is aimed at building a model which can accurately predict future financial distress of a company. The main goal leads in turn to several subgoals which will be described in this section.

Most of the research in the field of bankruptcy prediction has been carried out using the financial ratios of the company as the exploratory variables. In case of this study, availability of U.S. banks' financial data allowed us to increase the number of factors that may affect the future financial condition of the company. Also, previous work examined in their analysis only the financial results of the firm for the specific period before the predicted bankruptcy date. The novelty of this study is that it tries to find out if the changes in financial data during a specific period before expected bankruptcy can properly predict it. So, the first question raised in this study can be formally stated as:

- *Q1: How relevant are the changes in historical financial data for bankruptcy prediction?*

The number of features used in the analysis is quite significant. So, there is a need for dimensionality reduction and feature extraction in order to make the interpretation of the results amenable. In order to do it, two different techniques were implemented, namely: Principal Component Analysis (PCA) and Factor Analysis (FA). This leads to the second question:

- *Q2: Which dimensionality reduction and feature extraction technique is more useful in the analysis and, can we extract hidden information from the new features?*

[2] http://archive.fortune.com/2008/08/27/news/companies/demos$_b np.fortune/index.html$

The prediction accuracy varies depending on the technique used to build the model. Obtaining the best prediction accuracy is the main goal. So, different supervised machine learning techniques, namely, Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Random Forests and Supervised Self-Organizing Maps (XY-Fused networks) were compared to answer the following question:

- *Q3: Which machine learning technique best predicts the financial distress of a company?*

The main objective of the study is not only to predict the bankruptcy of a company based on the changes in financial results, but, also, to identify the historical time interval which best suits for predictions. So, the changes in financial data occurred during three different time intervals were considered. Those intervals are: year, half-year and quarter before prediction period. So, third question is stated as:

- *Q4: How far in advance can we predict bankruptcy?*

All the experiments and background information needed to answer the questions listed above will be disclosed in this thesis.

1.3 Outline

This study is divided into five chapters: introduction, background, materials and methods, classification experiments and conclusion. After this brief introduction, in chapter 2, we present the background of this work, including a brief presentation of the banking system of the United States of America; we discuss previous work in this area; and we describe the machine learning algorithms and concepts used in the analyses.

The first part of chapter 3 describes the data used to build the model and the data pre-processing steps that were conducted, including the dimensionality reduction and feature extraction processes.

The chapter 4 depicts the machine learning algorithms and their implementation with a focus on the technical side and the results of classification methods used for model development and, finally, the conclusions and some possible extensions of this work are presented in chapter 5.

Chapter 2

Background

This chapter provides an overview of the general information about the banking industry of United States of America and the main machine learning algorithms which were used in this work. Several related works in bankruptcy prediction were reviewed with comparative view of the techniques and approaches proposed by the authors.

2.1 Banking

Banking is an industry that works with the main means of economic exchange in the society. Banks offer different cash transactions, credits and a reliable place to keep savings for their clients, which includes savings accounts, checking accounts and deposits. Using the money that is stored in the deposits, banks can provide different types of loans that include mortgages, cars or business loans. Banking has always been one of the most vital parts of the United States economy, since most of its citizens have long-term liabilities to banks. By providing loans for families and entrepreneurs, banks provide an opportunity for people who do not have big savings to invest into the education, housing or businesses.

Insured by the Federal Deposit Insurance Corporation (FDIC), banks are the most reliable institution to deposit cash. Banks, in their turn, not only provide a safe storage for the cash, but also reward the depositors by a certain interest rate paid to them on monthly or yearly basis. According to FDIC requirements, banks are obliged to keep only 10 percent of their deposits and lend the other 90 percent with higher interest rates than they reward their depositors with. Therefore, it is easy for banks to make money. It is also true that banks should always balance it's operations in order to prevent financial problems in the future.

2.1.1 The banking system of the United States of America

The banking system of United States of America (U.S.) consists of a few different types of banking institutions, which are: commercial banks, community banks, retail banks, credit unions, mutual banks, savings banks, savings and loans, online banks and central banks (The Federal Reserve). Commercial banks provide different financial services to individuals and to businesses, while retail banking provides its services only to individuals and families. Community banks are smaller than commercial banks and provide fewer financial operations, concentrating on the local market. Credit unions and mutual banks are owned by individuals which are members of it. These types of institutions provide more personalized services. Savings banks are focused only on providing the saving accounts, while saving and loan institutions provide saving accounts and several types of loans. As can be seen from the name, online banks provide all of their services on-line and do not have physical locations.

A chain of events that included financial panic that damaged the U.S. economy in 19^{th} and the beginning of the 20^{th} century led to the creation of the Federal Reserve System or central bank. Among others, the central bank has four major functions, which are[1]:

- Monetary policy of the U.S., which includes conditions of credit and long term interest rates.
- Ensuring safety of the financial and banking system.
- Overlooking the stability and systemic risks of the financial system.
- Handling the payment system in both domestic and international domains.

2.1.2 Federal Deposit Insurance Corporation

The Federal Deposit Insurance Corporation (FDIC) is an independent agency of the U.S. government that protects the funds that depositors place in banks and savings associations. FDIC insurance is backed by the full faith and credit of the U.S. government. Since the FDIC was established in 1933, no depositor has lost a cent of FDIC-insured funds[2]. FDIC insurance covers all deposit accounts, including:

- Checking accounts
- Savings accounts
- Money market deposit accounts
- Certificates of deposit

FDIC insurance does not cover other financial products and services that banks may offer, such as stocks, bonds, mutual funds, life insurance policies, annuities or securities. The standard insurance amount is $250,000 per depositor, per insured bank, for each account ownership category[3].

2.1.3 Financial data

Financial data reflects the financial condition of a company and should be tied to a specific period of time in the past. It consists of day-to-day bookkeeping information. The financial data of a company is used in preparation of the company's financial statements. Financial statements are used by internal and external management to analyze and audit the business performance of the entity. In fact, there are four main types of financial statements:

- Balance sheets
- Income statements
- Cash flow statements
- Statement of shareholders' equity

[1]www.federalreserve.gov/faqs/about$_1$2594.htm
[2]https://www.fdic.gov/about/history/
[3]https://www.fdic.gov/deposit/deposits/faq.html

Balance sheets represent a company's financial position at the end of a specific time period. Observed periods of time can be a quarter, half year, year, or several years. Balance sheets express the fundamental equation:

$$Assets = Liabilities + Equity \tag{2.1}$$

Assets represent everything that an entity can own and include all intangible and tangible property. Tangible property is any physical property, whereas intangible property is non-physical, such as a patent or a goodwill.

Liabilities are everything that the company owes to others. Liabilities include such things as debt, accounts payable, wages, benefits, and taxes. Liabilities can be short-term, which means that the they will be due within a year, or long-term if the liabilities should be redeemed within a year or longer.

Stockholders' equity is known as the book value of a company. It has two main sources: the money initially invested in a firm and additional investments made later. Income statements summarize the total earnings during a period of time. The total earnings are calculated as the difference between total income and expenses. Cash flow statements disclose cash movements during the observed period. Statements of shareholders' equity show the changes in the owners' equity for the certain period.

2.2 Dimensionality reduction and feature extraction

The number of features in an input data vector could easily be as high as tens of thousands. Such a high dimensionality of data could have very adverse effect on data analysis and processing (Kung, 2014), this mainly affects the following factors:

- Computational cost. A large dimensionality usually leads to high computational complexity and power consumption both in the (off-line) learning and in the (online) prediction phases. The reduction of the number of dimensions helps to minimize the computation time and might prevent keeping irrelevant features in the data.

- Performance degradation due to sub-optimal search. A high data dimensionality may cause the numerical process to converge prematurely to a sub-optimal solution.

- Data visualization. It is connected with humans' inability to see objects geometrically in high-dimensional spaces.

- Data over-fitting. In supervised learning, when the input vector dimension far exceeds the number of training samples, data over-fitting becomes highly likely.

Dimensionality reduction offers an effective remedy to mitigate the above-mentioned problems. This data pre-processing step can also help in reducing the noise in the data, as it is possible to have irrelevant features or high multicollinearity in the data collected.

Next, we will discuss two primary techniques for dimension reduction and feature extraction considered in the thesis, namely Factor Analysis (FA) and Principal Component Analysis (PCA).

2.2.1 Factor Analysis

Factor analysis is a statistical technique that is widely used in psychology and the social sciences at large. It was originally developed by Spearman in 1904 in the area of human abilities in particular, to answer the question of why human abilities are always positively correlated. Factor analysis is a method for investigating whether variables of interest $X_1, X_2, \ldots, X_n$, are linearly represented by a smaller number of unobservable factors $F_1, F_2, \ldots, F_k$. The variables which make up these factors should be generally more correlated to each other than the factors are to each other.

The orthogonal model underlying FA can be described by Equation 2.2 (Hewson, 2009) :

$$X = \mu + \Gamma\alpha + \epsilon, \tag{2.2}$$

where X is an $1 * n$ random vector, μ represents a vector of unknown constants (mean values), Γ is an unknown $n * k$ matrix of constants referred to as the loadings, α is a $1 * k$ unobserved random vector referred to as the scores assumed to have mean 0, ϵ is $1 * n$ unobserved random error vector having mean 0 and by assumption a diagonal covariance θ referred to as the uniqueness or specific variance. Factor loadings (Γ) are defined as the correlations between variables and factors.

FA is an ambiguous model, as it is unchanged if we replace Γ by $K\Gamma$ for any orthogonal matrix K. This is a potential problem, but it can turn into an advantage because, with a reasonable choice of a suitable orthogonal matrix K, we can achieve a rotation that may yield a more interpretative result. FA therefore requires an additional stage, having fitted the model we may wish to consider rotation of the coefficients. We must keep in mind that orthogonally rotated factors have zero or negligible intercorrelation by definition. An oblique rotation provides a degree of correlation between factors to improve the mutual correlation between elements within the factors.

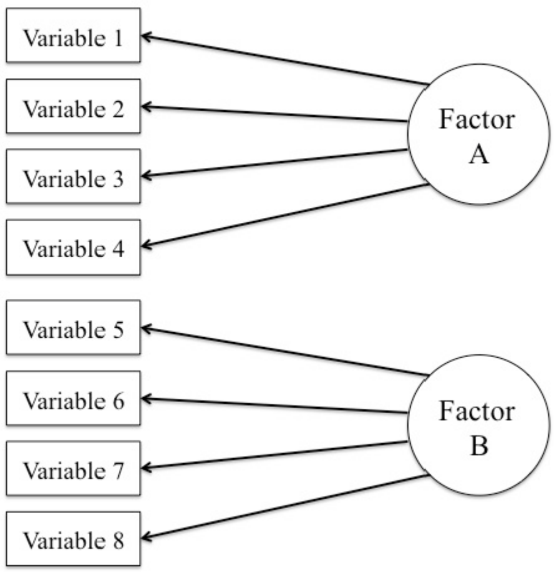


FIGURE 2.1: Conceptual overview of Exploratory Factor Analysis.

There are two main methods of FA:

- Exploratory factor analysis (EFA): attempts to uncover the nature of the constructs, influencing a set of responses.
- Confirmatory factor analysis (CFA): tests whether a specified set of constructs have the predicted effect on the set of responses.

The EFA is used to reduce the amount of data to be used or identify the number and nature of hidden latent factors in the data. The term "exploratory" is not used casually, as EFA does not test a model of factor structure, rather it examines the data set in search for statistically justified factors. It leads to subjectivity, since determining how many factors to select is a subjective and arbitrary process. (Plucker, 2003). The steps for performing EFA are as follows:

- Obtain the data
- Calculate correlation matrix
- Choose the number of factors for inclusion
- Extract initial set of factors
- Rotate the factors to obtain final solution
- Interpret factors structure
- Construct factor scores for further analysis
- Derive the new data set

Despite the many benefits, the FA has also some objections, listed below (Kline, 1994):

- Infinity mathematically equivalent solutions.
- The discrepancy of results. FA frequently leads to disagreement as to what are the most important factors in the problem.
- It is difficult to replicate FA. This statement comes from the first objection.

2.2.2 Principal Component Analysis

Principal Component Analysis is a dimensionality reduction technique of the feature extraction family. It is defined as an orthogonal transformation that linearly transforms a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components or PCs (Kung, 2014). PCA finds a linear projection of data into orthogonal basis system that has the minimum redundancy and preserves the maximum variance in the data. The projection process is illustrated in Figure 2.2 with an example of data set of observations x_n where $n = 1, .., P$, where x_n is a variable of dimensionality D. The goal is to project the data into a space of dimensionality $M < D$, while maximizing the variance of the projected input data (Bishop, 2006). In Figure 2.2, PCA seeks a space of lower dimensionality, which is denoted by the magenta line, such that the orthogonal projection of the data points (red dots) into this subspace maximizes the variance of the projected points (green dots). The new coordinates in the eigenvector basis, i.e. the orthogonal projections onto the eigenvectors, are the aforementioned PCs. The first

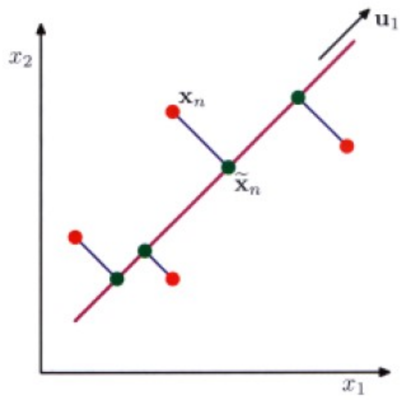


FIGURE 2.2: An orthogonal projection process in PCA (Bishop, 2006).

PC is chosen as a projection direction such that the projections of the data onto it have maximum variance.

PCA is an unsupervised approach, since it involves only a set of input features $X_1, X_2, \ldots, X_p$, and no associated response or target Y. The number of components extracted from PCA is equal to the number of observed variables in the analysis. The first PC identified accounts for most of the variance in the data. The second accounts for the second largest amount of variance in the data and is uncorrelated with the first principal component, and so on. The steps of the PCA procedure are as follows:

- Obtain data
- Standardize the data
- Calculate correlation matrix
- Calculate eigenvalues and eigenvectors of correlation matrix or perform Singular Vector Decomposition.
- Choose components to retain and form feature vectors
- Derive the new data set

The PCA model can be summarized by Equation 2.3, where X is a matrix of observed variables, Z is a matrix of scores on components and B is a matrix of eigenvectors.

$$X = ZB \tag{2.3}$$

The component score is a linear combination of observed variables weighted by eigenvectors. The scores of the first PC of a set of features $X_1, X_2, \ldots, X_p$ is the normalized linear combination of the features, which is presented in Equation 2.4 (G. et al., 2013).

$$Z_1 = b_{11}X_1 + b_{21}X_2 + \ldots + b_{p1}X_p \tag{2.4}$$

For dimensionality reduction purposes, the components accounting for maximal variance are retained while other components accounting for a trivial amount of variance are ignored. The eigenvalues indicate the amount of variance explained by each PC. Any PC with near-zero eigenvalues should be removed from analysis, as they do not explain a significant amount of variance in the data.

2.2.3 Factor Analysis versus Principal Components Analysis

FA and PCA are often confused with each other, it might occur due to the fact that there is one method of performing FA called Principal Component Extraction. These two methods should never be confused. PCA seeks orthogonal projections of the data according the variance maximization in order to achieve dimensionality reduction, and is not intended to find a reasonable interpretation for all the components that are kept. FA, instead, tries to study the covariance (or correlation) relationships between many variables and is based on creates unobservable or latent random quantities called factors (Hewson, 2009) and it is, this, a latent variable model. So, PCA observes the relationships between the individual and total (common and error) variances shared between items, while FA observes the relationship between the individual item variances and common variances shared between items. Sometimes, in the early stages of an analysis, FA is preferable to PCA, as it allows you to measure the ratio of an item's unique variance to it's shared variance, known as its communality.

FA and PCA have two main conditions. Firstly, there must be some relation/connection (correlation) between the variables. In addition, the larger the sample size, especially in relation to the number of variables, the more reliable the resulting factors. The sample size is less important for factor analysis, since the communalities of objects with other objects are high or relatively high. However, PCA or FA should never be performed if the number of variables is greater than the number of observations in the data set.

There are several similarities between the FA and PCA techniques (Klinke, Mihoci, and Hardle, 2010; Hardle and Simar, 2003), including:

- Based on a linear model
- Aim to reduce the number of data features
- Can be used on covariance or correlation matrix
- Provide similar results for the resulting PCs and latent factors (without rotation).

And at the same time they have differences, as listed below:

- PCA has an importance ranking of the components determined by the eigenvalues while in EFA factors are all equal in the analysis and only might be ranked after rotation based on interpretability of factors which is nonobjective.
- PCA is based on a well-defined algorithm, whilst the fitting factor analysis model includes many numerical procedures. The non-uniqueness of the factor analysis procedure opens the door for subjective interpretation and, therefore, produces a range of different results.
- PCA optimises the total variance. Since the total variance is the sum of squared distances to the data centre it is obvious that the covariance or correlation structure of the data does not play any role. EFA aims to reproduce the covariance or correlation matrix as well as possible.

2.2.4 Selection of the number of components or factors to retain

There are several techniques for the identification of the number of dimensions that should be kept for further analysis, and most of them are commonly used both in FA and PCA. Several of these methods will now be discussed further in this section.

The first one is the Kaiser-Guttman rule, also referred to as "the Kaiser criterion", or "the eigenvalues > 1.0 rule". It proposes dropping factors whose eigenvalues are less than one, since the variance explained by each of these factors is less than the variance explained by a single variable. The Kaiser–Guttman rule is widely used because of its simplicity.

The second approach, called Cattell scree test, is a graph-based *ad hoc* technique that uses the eigenvalues that are taken from the input or reduced correlation matrix. As shown in Figure 2.3, the eigenvalues form the vertical axis and the factors form the horizontal axis. The graph is inspected to determine the last substantial decline in the magnitude of the eigenvalues or the point where eigenvalues substantially change slope. This graph is the most popular method for determining the number of factors/dimensions. A limitation of this approach is that the results of the scree test might be ambiguous and open to subjective interpretation (Brown, 2006).

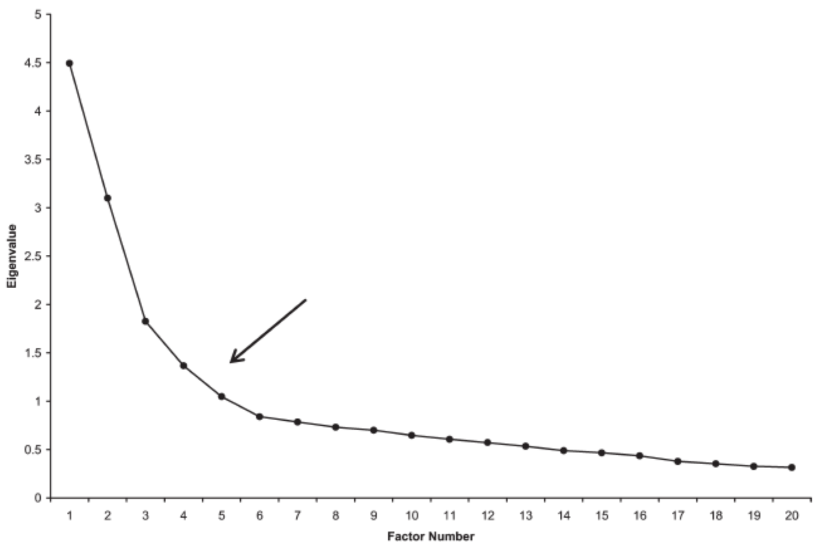


FIGURE 2.3: Scree test of eigenvalues. Arrow indicates region of the curve where the slope changes.

Another criterion is to set a certain percentage of variance that needs to be explained, and then keep enough factors/dimensions to achieve this variance from data. Usually, the lower limit is set to be at least 50% of the total variance in the data.

The choice of the wright technique among all the above mentioned methods is not straightforward. Perhaps the best advise is to use the number of factors that best agrees with the goals of analysis. If there is no need of substantial dimensionality reduction, it might be better to keep most, if not all, of the factors/dimensions during the early stages of the analysis. If the goal is to reduce the dimensionality of data, then it is best to keep enough factors/dimensions so that they will explain a reasonably large percentage of the variation in data.

2.3 Supervised Learning

Most machine learning problems can be considered to fall into one of three categories: unsupervised, semi-supervised or supervised. The previous part of this chapter discussed unsupervised methods in this work, and the next part of the thesis is devoted to supervised learning techniques.

The main characteristics of supervised learning models is that for each observation of the predictor measurement x_i, $i = 1, ..., n$ there is an associated response measurement Y_i, and the goal is fitting a model that relates the response to the predictors. This model should accurately predict the response for future observations or better understand the relationship between the response and the predictors (G. et al., 2013). All of the machine learning classification methods discussed and used in this thesis are examples of supervised learning.

2.3.1 Support Vector Machine

The Support Vector Machine (SVM) is a supervised learning algorithm which is used for classification and regression analysis. It is an extension of the support vector classifier that results from expanding the feature space in a specific way using kernels. The SVM constructs a hyper-plane, or set of hyper-planes, in a high- or infinite-dimensional space, which can be used for classification. The best separation is achieved by the hyper-plane that has the largest distance to the nearest training data point of any class, since in general, the larger the margin, the lower the generalization error of the classifier. SVM is good at classification because it minimizes the generalization error rather than the training error (Vapnik, 1998).

In the case of non-linear SVM, the objective is to linearly divide the data in a higher-dimensional space. This is done via a kernel function, in what is known as the "kernel trick", which has its own set of parameters. When it is translated back to the original feature space, the result is non-linear. The number of support vectors, found for each model, depends on how much slack is allowed, but it also depends on the complexity of the model. Each "twist and turn" in the final model in the input space requires one or more support vectors to define. Ultimately, the output of an SVM is the support vectors and an alpha parameter, which, in essence, is defining how much influence that specific support vector has on the final decision. In the case of non-linear SVM, accuracy depends on the trade-off between a high-complexity model which may overfit the data and a large-margin which will incorrectly classify some of the training data in the interest of better generalization. The number of support vectors can range from very few to every single data point if we completely over-fit our data. This trade-off is controlled via parameter C and through the choice of kernel and kernel parameters. The computational complexity of the model is linear in the number of support vectors. Fewer support vectors mean faster classification of test points.

SVM is, primarily, a non-parametric method, yet as previously mentioned, some hyperparameters do need to be tuned before optimization. In the Gaussian kernel case, there are two hyperparameters: C, which is the penalty term and σ, the width of the exponential. A too small value of σ causes $k(xi, xj) = 0.$, $ij.$, i.e., each sample is considered as an individual "cluster". While a too high value causes $k(xi, xj) = 1.$, i.e., all samples are considered neighbours. Thus, only one cluster can be identified. The choice of σ should reflect the range of the variables, to be able to detect samples that belong to the same cluster from those that belong to others clusters. This is usually done by a cross-validation step, where several values are tested (Kung, 2014).

2.3.2 K-Nearest Neighbors

The K-nearest neighbors (KNN) classifier is aimed at estimating the conditional distribution of Y given X, and then classify a given observation to the class with highest estimated probability. Given a positive integer K and a test observation x_0, the KNN classifier first identifies the neighboring K points in the training data that are closest to x_0, represented by N, and then estimates the conditional probability for class j as the fraction of points in N whose response values equal j (G. et al., 2013):

$$Pr(Y = j|X = x_0) = \frac{1}{K} \sum_{i \in N} I(y_i = j) \tag{2.5}$$

KNN applies Bayes' rule and classifies the test observation x_0 to the class with the largest probability.

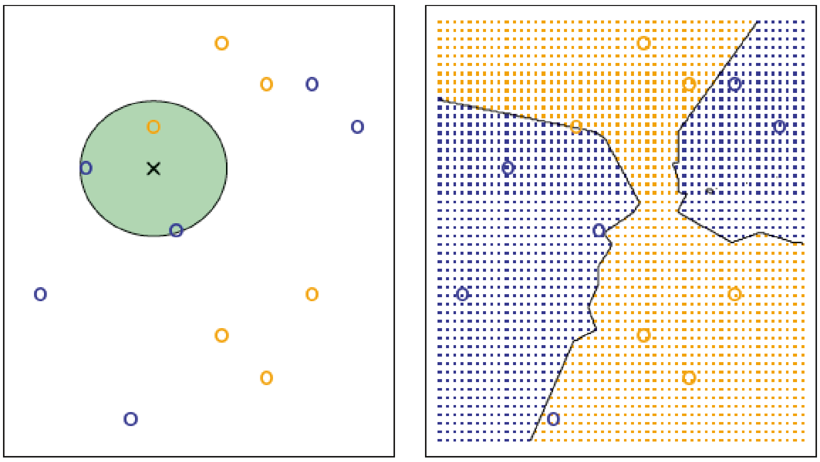


FIGURE 2.4: The KNN approach, for $K = 3$ (G. et al., 2013).

Figure 2.5 provides an illustrative explanation of the KNN approach. A set of training data consisting of blue and orange class observations is presented on the left side of the plot. The purpose of this problem is to predict the class (blue or orange) of the black cross labeled data point, given that K is equal to three. KNN first finds the three observations that are closest to the cross. These neighborhood of three nearest observations is marked by the circle. It consists of two blue points and one orange point, with resulting estimated probability of $2/3$ for the blue class and $1/3$ for the orange class. Therefore, KNN will predict that the black cross belongs to the blue class. The right-hand side of Figure 2.5 illustrates the result after application of KNN approach at all of the possible values for $X1$ and $X2$ with drawn KNN decision boundary. The choice of K strongly influences the resulting KNN classifier. We can obtain the optimal model by varying the K value and comparing the training and validation errors. Although this is an overall very simple approach, KNN can often produce surprisingly good results. The summary of some strong and weak points (Lantz, 2013) of the KNN algorithm are presented below .

Strengths:

- Simple and effective
- Makes no assumptions about the data distribution
- Fast training phase

Weaknesses:

- Slow classification phase
- Does not produce a model as such, which limits the ability to find relationships among features
- Requires a large amount of memory

2.3.3 Random Forests

Bagging or bootstrap aggregation is a method to reduce the variance of the estimated prediction function. Bagging seems to be particularly well suited for for high-variance and low-bias procedures, such as trees. In the case of regression problems, we should repeatedly fit the same regression tree to bootstrap sampled versions of the training data, and average the results. For classification, each time a committee of trees votes for the predicted class.

Random Forests (RF) is a machine learning algorithm that take account of all the features at the same time. It focuses only on ensembles of decision trees. RF is a modification of bagging that builds a large collection of de-correlated trees, and then averages them. This method was developed by Leo Breiman and Adele Cutler. The algorithm builds *ntree* trees repeating the following steps (Usuelli, 2014):

- step1: Subset the data to build the tree by choosing a random row from the data *sampsize* times. Each row can be chosen more than once and in the end we have a table with *sampsize* random rows.
- step2: Randomly select a *mtry* number of features
- step3: Build a decision tree based on the sampled data

RF combine versatility and power in a single machine learning approach. Given that an ensemble uses only a small random portion of the full set of features, it can process very big data sets, where such dimensionality might lead to other models failing. Nevertheless, error rates for most learning tasks are almost as good as for any other method. The strengths and weaknesses (Lantz, 2013) of the model are summarized below.

Strengths:

- Selects only the most important features
- Can be used on data with an extremely large number of features or observations
- Performs well on most problems

Weaknesses:

- It may require some effort to tune the model to the data
- The model is not easily interpretable

2.3.4 Supervised Self-Organizing Map

Self-Organizing Maps (SOM) are an effective tool for visualization of high-dimensional data. The SOM (also known as Kohonen maps) algorithm was invented by Professor Teuvo Kohonen back in 1982, aiming to define a neuro-computational bioplausible model. The SOM produces a nonlinear, ordered, smooth mapping of a high-dimensional data on a regular, low-dimensional (usually 2D) grid (Kohonen, 2001).

The SOM consists of an input level that distributes input data to each node (neuron) at the second level, the so-called competitive level. Each of the nodes in the second layer acts as an output node. Each node in the competitive layer is connected to other nodes in its neighbourhood. Neurons in the competitive layer have strong connections with the nearest neighbors and weak connections with more distant neurons. Each node i in the map has a weight vector w_i and the number of elements in the weight vector is equal to the number of features in input vector. Each node (neuron) i is defined by a position in a pre-defined grid of fixed dimension. The SOM is continuously updated during the training phase by randomly choosing one input example x_k and applying the following algorithm (Buessler, Urban, and Gresser, 2002; Almendra and Enachescu, 2014):

- Choose the winning unit i^* that minimizes the distance $||x_k - w_i||$
- The weights of units are updated according to following formula:

$$w_i = w_i + \rho\Phi(i, i^*)(x_k - w_i), \tag{2.6}$$

where ρ is the learning rate, $\Phi()$ is the neighboring function, which is a monotonically decreasing function of the distance between units i and i^*. According to this algorithm, the weight vectors of the winner node and its neighbors are updated and, thus, they become more similar to the input vector, while this similarity will decrease for more distant neurons. The weight correction process is repeated iteratively until all vectors in the training set are presented a sufficient number of times to the network. Nowadays, SOM is most commonly used in the areas of data mining, in particular, data visualization, clustering in biomedical analysis, engineering sciences, macroeconomics and finance.

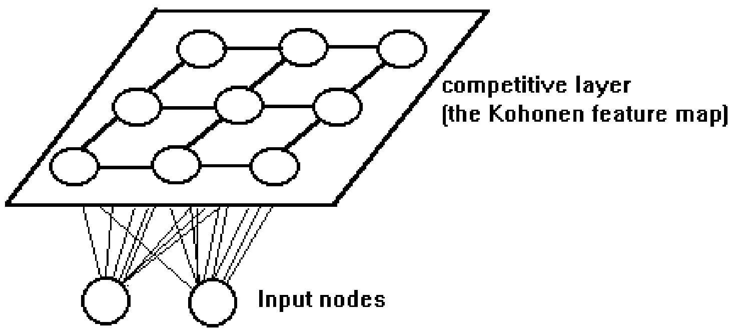


FIGURE 2.5: Architecture of the Self-Organizing Map.

The SOM is an unsupervised technique, but there are several supervised variants of SOM such as counter propagation artificial neural network (CP-ANN), supervised Kohonen networks (SKN) and XY-Fused networks (XYF). The Supervised Self-Organizing Map (SSOM) can be used for classification problem, where $Y =$

$(Y_1, Y_2, ..., Y_C)$ is a target output with C classes, $X = (X_1, X_2, ..., X_N)$ is an input vector with N number of features. Y represents a binary vector containing class information, where only the class index to which it belongs is set to 1. The difference between SOM and SSOM is that an additional vector of class information is included in the training and it introduces an additional factor that organizes the map. This data model allows class information to influence the topological ordering of the map during training process. Then, the trained map is used for predicting the unknown Y dimension. The extent to which class information affects map can be controlled by class weight, which can be adjusted depending on how class information is used to train the map: a low value causes the map to be close to unsupervised, and a high value may overfit the data (Xiao et al., 2006; Wongravee et al., 2010).

The architecture of the SSOM model for classification problems is a three-layer neural network, as shown in Figure 2.6. The first layer is the input layer which consists of N nodes (neurons) corresponding to the number of features in the input data. The second layer nodes are constructed during training phase and each node represents a reference pattern. The third layer is the output where each node represents a specific class. Each node in the second layer is connected with the first layer nodes through connections w_{ji} . The weight vector w_j of the dimension N represents the reference pattern of the j-th node in the cluster layer. When the model obtains the input and associated target output, the input vector X_i is transmitted to the cluster layer, and each node in the cluster layer then calculates the degree to which the input vector X_i belongs to cluster j. Next, the system makes a cluster choice by selecting the winning node j with maximum choice function value from all the nodes j in the cluster layer. If the winning node j belongs to the correct class defined by the target output vector, the weight vector of the winning node and those of its neighboring nodes whose classes are the same as the winning node j will be updated. However, if the winning node does not represent the class to which it belongs, the system will search for the next best cluster node $j*$ whose class is the same as the target output (Thammano and Kiatwuthiamorn, 2007).

The explanation of the SSOM architecture above explains general model, but there might be some differences in different types of SSOM. In CP-ANN, the winning node on the input layer determines the position of the winning node on the output layer, so, the output layer of the simplified CPN model is developed exclusively by the topology present in the input space. Therefore, the CPN model cannot be considered as a truly supervised method. During the training of the SKN model, the input and output layers must be glued and training process works like in a standard SOM, but the information in the output layer is used to indicate the winning node in the learning phase. The main disadvantage of the SKN network is that the user must determine the right balance between input and output objects. Correct scaling of input and output vectors has a huge importance on model creation. Imbalance in inputs and outputs can negatively affect the model efficiency. In XY-Fused networks, the fused similarity is calculated from both input and output layer and is used to determine the position of the winner node. The set of similarities obtained for an object X and the input map units is combined with the similarities corresponding to the output object Y and the output map such that common winning unit for both maps is determined (Vasighi and Kompany-Zareh, 2013; Melssen, Wehrens, and Buydens, 2006).

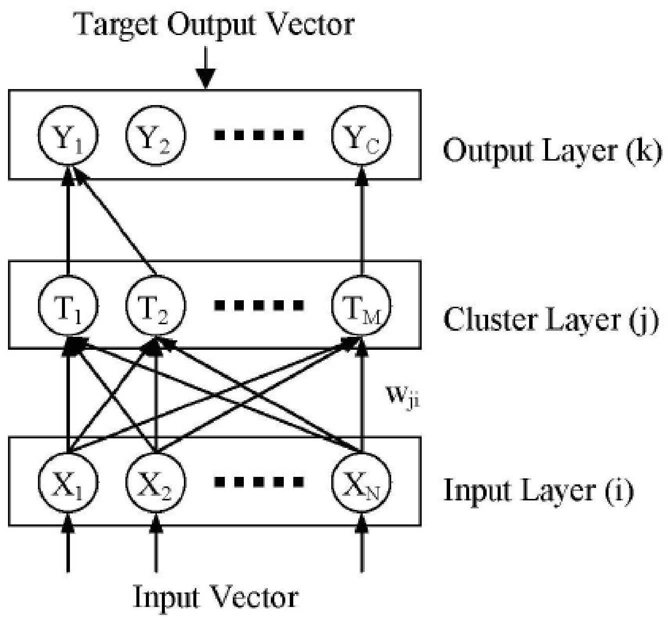


FIGURE 2.6: Architecture of the Supervised SOM for classification problem (Thammano and Kiatwuthiamorn, 2007).

2.3.5 Cross Validation

The data analysis might face a bias caused by the particular sample chosen. Each class in the data set should be represented in equal proportions in the training and testing sets. If all examples with a certain class were omitted in the training set, the classifier extracted from this data will not work well with examples from this class. A simple way to prevent it, is to use statistical technique called cross-validation. In cross-validation, a fixed number of folds (k) of the data should be chosen. Then, the training sample is divided into k subsets, each of which has the same number of samples. The classifier is trained k times, in each iteration the one of subset is used for testing and the remaining data is used for training. For example, in the i-th iteration ($i = 1, ..., k$) iteration the classifier is trained on all subsets except the i-th one, then, the classification error is computed for the i-th subset. The procedure should be repeated k times so that in the end, every instance has been used exactly once for testing. This is called k-fold cross-validation. Different k-fold cross-validation experiments with the same learning scheme and data might provide different results due to the effect of random variation in choosing the folds (Witten, Eibe, and Hall, 2011).

2.4 Review of related works

Financial distress prediction is a well-studied topic. There is no standardized procedure to access companies' full internal data, so, most models proposed in the literature rely on only main financial ratios which are easy to obtain as public companies are bound to disclose their main financial results. There are some rare cases though in which not only financial indicators are used. For instance, Ptak-Chmielewska and Matuszyk (2018) in their work "The importance of financial and non-financial ratios in SMEs bankruptcy prediction" (Ptak-Chmielewska and Matuszyk, 2018) used financial and non-financial data in their analyses. The usage of non-financial data together with financial data improved the results of models which were previously based on only financial indicators. In addition, some researches are trying to determine whether the financial results of bankrupt firms differ based on demographic data. Lukason (2012) used Independent Samples Median Test to check whether medians of different pre-bankruptcy financial results changes vary through firm types (Lukason, 2012). Based on the data of Estonian bankrupt firms for the period 2002-2009, it was proved, that there are a differences in the financial indicators for different industries, size groups, bankruptcy years, insolvency types and varying levels of control.

According to my review of the literature, some studies use dimensionality reduction or feature extraction techniques as pre-processing step. The study conducted, *inter alia*, by Adalessossi (2015) used PCA to explore hidden relationships between variables (Adalessossi, 2015). Chen (2011) used PCA for dimensionality reduction in the study "Bankruptcy prediction in firms with statistical and intelligent techniques and a comparison of evolutionary computation approaches". It was identified that with nearly 80% fewer financial ratios, the prediction performance is still able to provide highly-accurate forecasts of financial bankruptcy (Chen, 2011). Vaclav and Hampel (2016) used filter based feature selection algorithms like Gain ratio, Chi-square and Relief in order to obtain attributes with the best information value(Václav Klepáč, 2016). A SOM model was used by Kiviluoto (1998) in the paper "Predicting bankruptcies with the self-organizing map" as an exploratory preprocessing step to visualize the differences between companies that go bankrupt and those that do not (Kiviluoto, 1998). Arora and Saini (2014) applied Independent Component Analysis (ICA) on the input data set comprising financial ratios to choose the most significant to be considered as input to the further analysis (Arora and Saini, 2014).

Plenty of techniques have been used in bankruptcy prediction. We will discuss most popular algorithms in the remaining part of this section. SVM is one of the most frequently used classification techniques in the area of bankruptcy prediction. The studies conducted by Chen (2011) and Kalyan and Amulyashr (2015) analyzed bankruptcy prediction with different machine learning techniques like logistic regression, decision trees, RF, Naive Bayes, neural networks and SVM, the results showed that SVM outperformed other techniques (Chen, 2011; Kalyan and Amulyashree, 2015). The authors of the paper "Prediction of Bankruptcy with SVM Classifiers Among Retail Business Companies in EU", Vaclav and Klepac (2016), applied the SVM method with linear, polynomial and radial kernels to obtain the best bankruptcy prediction results (Václav Klepáč, 2016). The data used in the study consists of financial data of 850 medium-sized retail business companies in EU from which 48 companies were bankrupt in 2014. One of the questions raised in this paper is whether it is possible to predict bankruptcy 1–5 years before the bankruptcy time. The results indicated that the longest prior-to-bankruptcy period models are not efficient enough to predict the bankruptcy. The SVM classifier based on RBF kernel performed best according to accuracy for 1 year-ahead prediction.

Hauser and Booth (2011) investigated the accuracy of bankruptcy prediction using financial ratio data of U.S. firms from 2006 till 2007 (Hauser and Booth, 2011). They compared the results of robust logistic regression with the Bianco and Yohai (BY) estimator versus maximum likelihood (ML) logistic regression and BY. With both the 2006 and 2007 data, BY robust logistic regression improved the classification results of ML logistic regression in the training and testing sets. The study "A Cash Flow Based Model of Corporate Bankruptcy in Australia", by Jones (2016), employs binary logistic regression to predict corporate bankruptcies in Australia using cash flow based ratios (Jones, 2016). The results outperformed a logit model estimated on Altman Z-score variables.

The Z-score formula was devised in 1968 by Edward I. Altman. This formula is used to predict the probability that a firm will face financial distress. The advantages of this method are that it is easy to calculate and provides quite satisfactory results. Craciun and co-workers (2013) tested the suitability of Altman's model to predict the financial health of Romanian companies in the period of financial crisis (Crăciun et al., 2013). The data used in the study included financial ratios of 60 Romanian companies for the period between 2005 and 2009. Altman's model obtained a satisfying result for the economic period in which this model was developed (1946-1965), however, it failed predicting the bankruptcy of Romanian firms under an unstable economic environment. Adalessossi (2015) applied Altman's Z-scores to predict the probability of bankruptcy of West African's firms using the financial statements for 2013. The analysis overall provided fair results (Adalessossi, 2015).

Artificial neural networks have performed well in business-related classification problems including bankruptcy prediction. Arora and Saini (2014) used Fuzzy SVMs to predict financial distress in companies in "Bankruptcy Prediction of Financially Distressed Companies using Independent Component Analysis and Fuzzy Support Vector Machines". Surprisingly, Fuzzy SVMs yielded an accuracy of around 94%. Lately, The SOM method has became more popular in classification problems like bankruptcy prediction. Back, Oosterom and Sere (1994) examined the prediction power of the SOM algorithm, the backpropagation network and the Boltzmann Machine. The resultas showed that the backpropagation net performs best in bankruptcy prediction (Back, Oosterom, and Sere, 1994). Serrano-Cinca (1996) describes the usage of SOM for financial health analysis in his work named "Self organizing neural networks for financial diagnosis". This model was used separately as well as in combination with other models like Linear Discriminant Analysis and a Multilayer Perceptron artificial neural network. According to Serrano-Cinca, the flexibility of the neural model for combining and adapting to other structures, whether neural or otherwise, guaranteed a bright future for this type of model (Serrano-Cinca, 1996). Kiviluoto (1998) utilized the SOM algorithm in qualitative analysis to visually examine difference between bankrupt and non-bankrupt firms and in the classification analysis as a vector quantizer, to predict financial distress in firms (Kiviluoto, 1998).

In this thesis, several machine learning and related algorithms, namely KNN, RF, SVM and supervised SOM will be used to predict the financial distress of U.S. banks for the period 1993-2017. The prediction will be made based on the results of dimensionality reduction and feature extraction techniques like PCA and EFA, which will be obtained from the features corresponding to the changes in the financial results of U.S. banks based on three different time periods: a quarter, half-year and a year before bankruptcy.

Chapter 3

Materials and Methods

In this chapter, the data sources are summarily described, followed by the data preprocessing methods and the strategies to increase the interpretability of the data.

3.1 Data Source

The data used in this study was retrieved from the Federal Deposit Insurance Corporation (FDIC) database. The FDIC provides the list of U.S. insured banks which went bankrupt during the period 1992-2017. Also, financial organizations insured in FDIC submit quarterly reports with financial results, which are publicly available on the FDIC website. According to FDIC, the total number of the U.S. banks which went bankrupt between 1992-2017 reached 845.

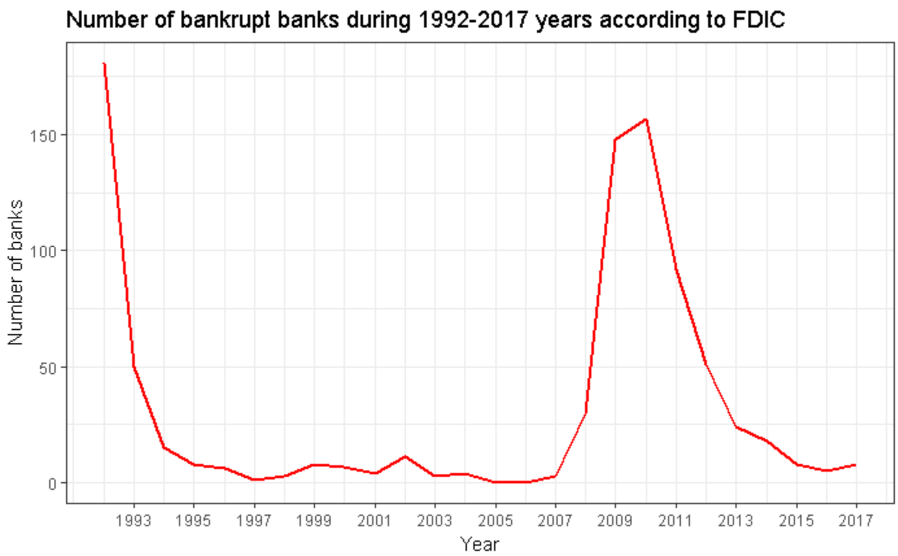


FIGURE 3.1: Number of banks which faced financial distress during the 1992-2017 period in the U.S.

All quarterly reports with financial data of banks for the 1992-2016 period were extracted from the FDIC website to be used in the prediction of the financial distress of observed banks during 1994-2017 period. The data for each quarter consists of up to 60 financial reports in CSV format. The number of banks listed in the reports varies from 5,679 up to 13,973.

After merging all financial results for each quarter, we obtained overall 1,034 financial indicators common to all observed time periods. These indicators were used to form exploratory variables by calculating the percentage changes of each

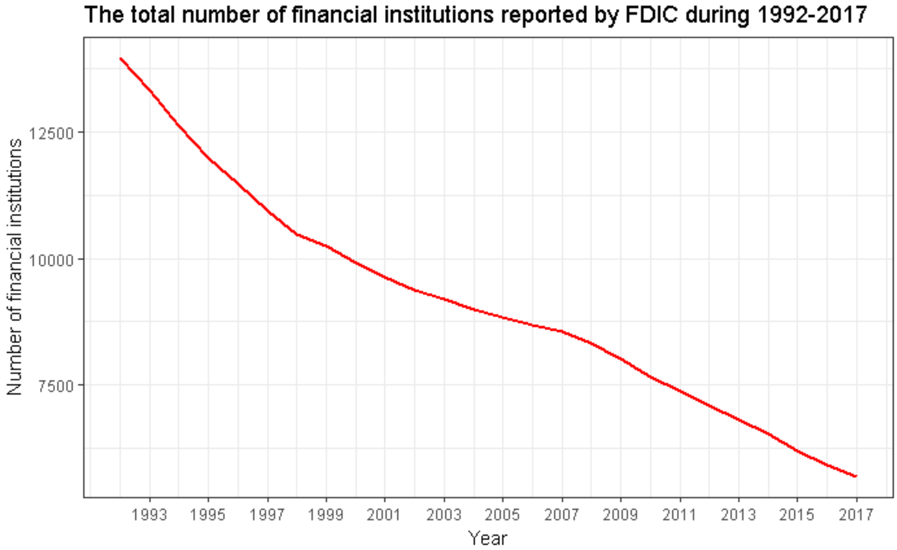


FIGURE 3.2: Total number of banks in U.S. during 1992-2017

financial indicator during one of the three proposed time intervals; therefore, three different data sets with 1,034 features were created.

The final data sample consists of changes in financial results of all bankrupt banks and randomly selected non-bankrupt banks. The formula for computation of changes in financial results is presented below:

$$Fr_{ij} = \frac{F_{i(j-1)} - F_{i(j-2)}}{F_{i(j-2)}}, \tag{3.1}$$

where, Fr_{ij} - changes in financial results of bank i for the period j; $F_{i(j-1)}$ - financial results of bank i for the period prior to j period; $F_{i(j-2)}$ – financial results of bank i for the period prior to $j-1$ period. The periods of calculation might be the following: 1) quarter-based for scenario 1, 2) half-year-based for scenario 2, 1) year-based for scenario 3.

3.2 Data Pre-processing

The data samples for our analyses were built from all bankrupt and randomly selected non-bankrupt banks' financial results. Non-bankrupt banks' data was selected from the same period that for bankrupt banks, and the proportion of bankrupt-to-non-bankrupt is equal to 1/5. Given the requirement that the financial history of the enterprise be known well enough, if there was no data available for an observed period before the bankruptcy date of the company, the bank was excluded from the analyzed sample. As mentioned, the data sets for all three scenarios consist of 1,034 exploratory variables, including the demographic data.

In the data pre-processing step, the main concern was to check if the data set had any missing values. The final data sets were checked for missing values and all variables with more than 90% of missing values were removed from them. The remaining missing values, even if this is not the optimal procedure, were replaced with "0" values. After removal, the total number of variables decreased to only

TABLE 3.1: Groups of exploratory variables.

Group name	Number of variables
Assets and Liabilities	30
Performance and Condition Ratios	27
Total Deposits	13
Income and Expense	12
Net Loans and Leases	9
Securities	3
Changes in Bank Equity Capital	2
Total Interest Income	2
1-4 Family Residential Net Loans and Leases	1
Additional Noninterest Expense	1
Cash and Balances Due	1
Maturity & Repricing for Loans and Leases	1
Nontransaction Accounts	1
Time Deposits at the $100,000 Threshold	1
Total Interest Expense	1
Transaction Accounts	1
Total	106

139 in all scenarios. As demographic data is not used in further analysis, the corresponding columns were also removed. In the end, the data set was composed of 106 exploratory variables and 1 response variable. The response variable is categorical, where the value "1" refers to bankrupt entities, and "0" corresponds to non-bankrupt entities.

In order to simplify the visualization of the data, exploratory variables were split into 16 groups based on the definition provided by FDIC reports. The groups of data are presented in full in Table 3.1. The full list of explanatory variables for each scenario with the corresponding groups are presented in Appendix A, Appendix B and Appendix C. As some of groups of data were subgroups of other groups, the variables were merged into 6 "super-groups" only for visualization purposes.

The data sets were checked for multi-collinearity, and the "Variance Inflation Factors" (VIF) test detected very strong multi-collinearity in all three scenarios.

3.2.1 Scenario 1: Quarterly changes in financial results

This data set was built by computing the percentage of change in financial results between 1 and 2 quarters prior to the possible bankruptcy date.

The data from FDIC reports from 4Q 1992 till 3Q 2017 were used to forecast bankruptcy for the period from 2Q 1993 till 4Q 2017.

The data set consists from 399 observations of bankrupt banks and 1,995 observations of non-bankrupt banks. The original total number of variables including demographic data is equal to 1,034. After the removal of variables with too many missing values, the total number of variables decreased to 139. The final data set after demographic data removal consisted of 106 exploratory variables and 1 response variable. The data set can be considered as medium-sized, consisting of just 2,394 entries.

The boxplots of variables from the "Loans and Leases" group, broken down by class value, are presented in Figure 3.3. The boxplots of the remaining groups are

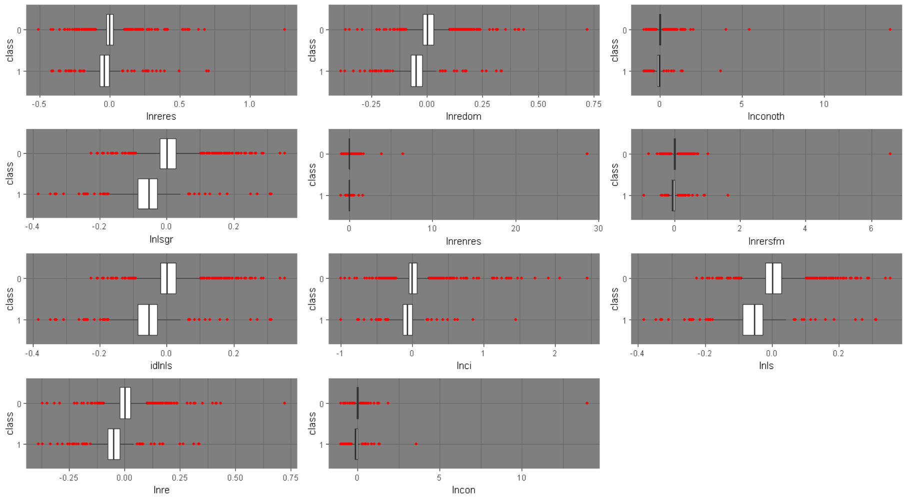


FIGURE 3.3: *Loans and Leases* group boxplots: Scenario 1

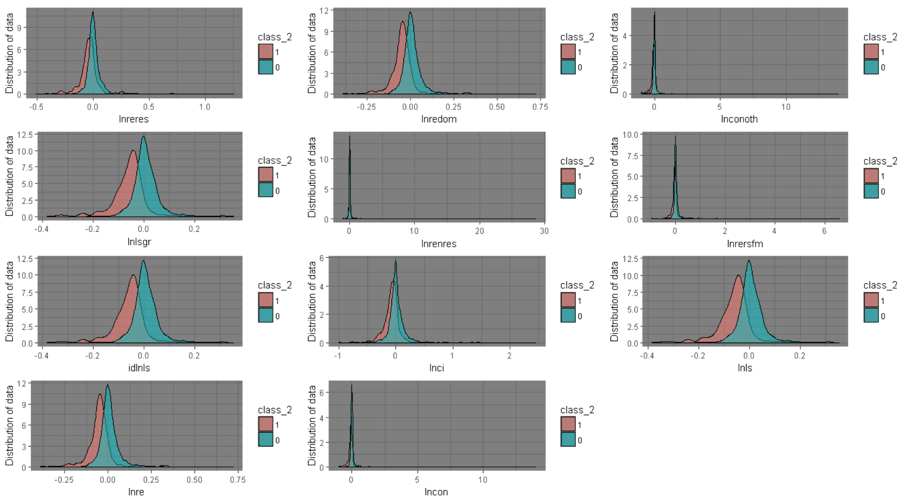


FIGURE 3.4: *Loans and Leases* group density plots: Scenario 1

presented in Appendix D to avoid cluttering the document. There is a common pattern for some of the variables: the bankrupt banks' exploratory mean values of the variables are smaller than zero, which means negative percentage change in financial results, while the mean values of exploratory variables of non-bankrupt entities are close to 0 or slightly over 0.

The density plots of the variables were broken down by class value. It helped to understand the overlap of classes for any given attribute. Figure 3.4 shows the density plots of the "Loans and Leases" group variables. The red colored part of the density plot presents bankrupt entities. As we can see, the density of bankrupt and non-bankrupt entities are almost equal. Also, we should take into account that the number of bankrupt banks is 5 times lower than the non-bankrupt banks in the data set. In the density plots, we can see that those of bankrupt banks are shifted to the left, as related to the non-bankrupt banks. The density plots of the remaining groups of variables are presented in Appendix E for further detail.

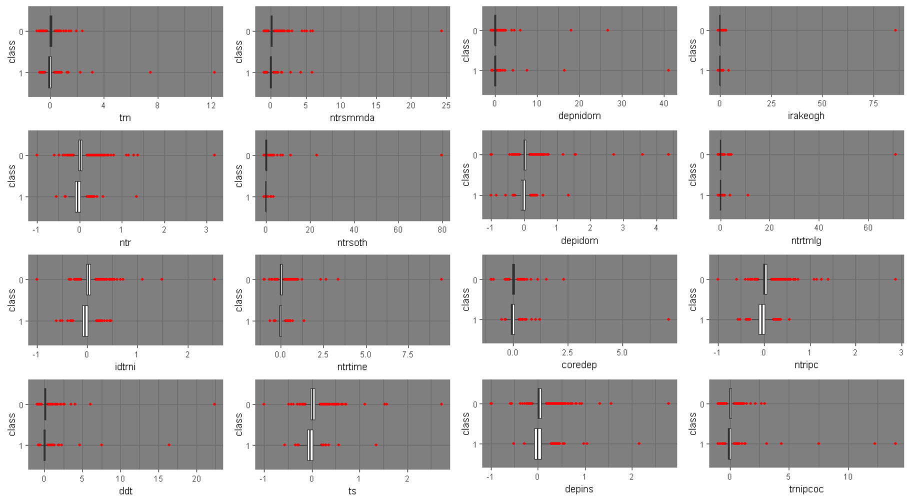


FIGURE 3.5: *Total Deposits* group boxplots: Scenario 2

3.2.2 Scenario 2: Half-year changes in financial results

The data set for this scenario was built from the percentage changes in financial results between half-year and 1 year prior to the possible bankruptcy date. The data from FDIC reports from 4Q 1992 till 2Q 2017 were used to forecast bankruptcy for the period from 4Q 1993 till 4Q 2017.

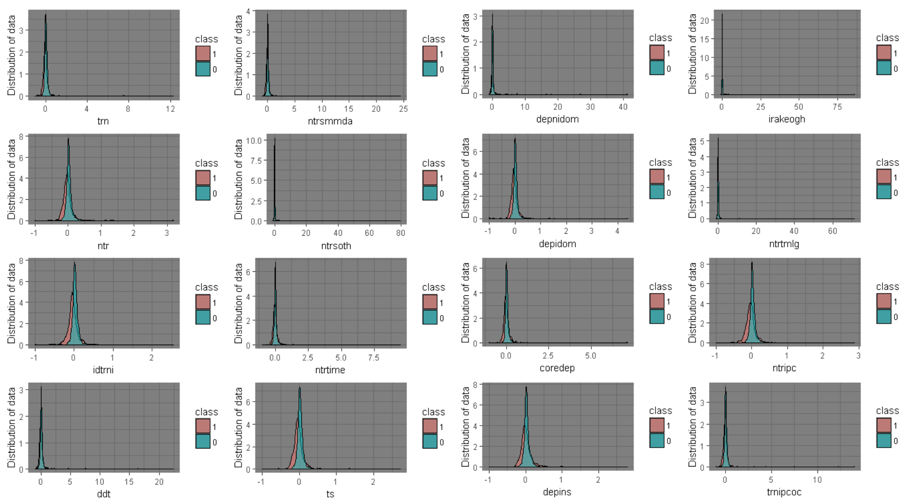


FIGURE 3.6: *Total Deposits* group density plots: Scenario 2

If there was no data available for 1 year before the bankruptcy date of the company, the bank was excluded from the sample. The data set consisted of 390 observations of bankrupt and 1,950 non-bankrupt banks' results, with an overall 2,340 entries.

The boxplots of variables from the "Total Deposits" group, broken down by class value are presented in Figure 3.5, remaining boxplots are presented in Appendix F. The density plot was also broken down by class value. Figure 3.6 shows the density plots of the "Total Deposits" group variables. The red colored part of the plot represents bankrupt banks. The density of bankrupt and non-bankrupt entities are almost equal except that the density plots of bankrupt banks are shifted to the left related to the non-bankrupt banks. The density plots of remaining groups of variables are presented in Appendix G.

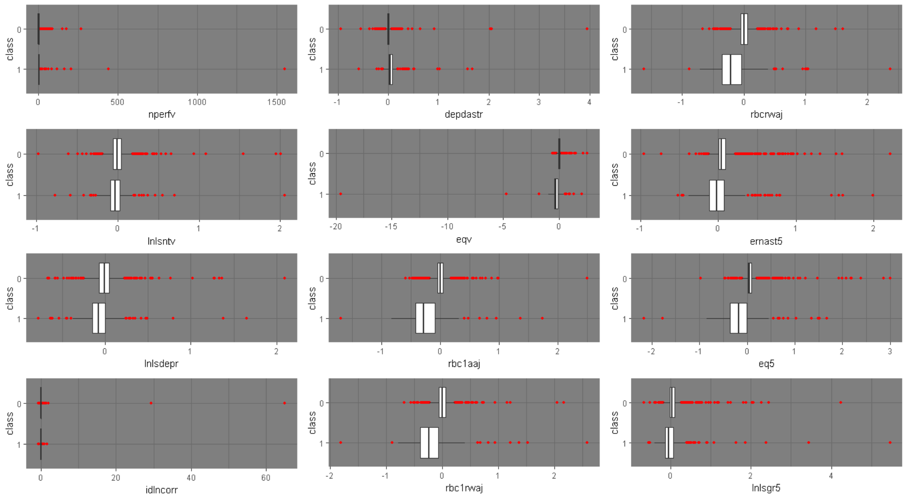


FIGURE 3.7: *Performance and Condition Ratios* group boxplots: Scenario 3

3.2.3 Scenario 3: Yearly changes in financial results

The data set for this scenario represents the percentage changes in financial results between 1 year and 2 years before possible bankruptcy date. The data from FDIC reports from 4Q 1992 till 4Q 2016 were used to forecast bankruptcy for the period from 4Q 1994 till 4Q 2017. The data set includes 375 entries of bankrupt banks and 1,875 of non-bankrupt banks, which is 2,250 entries in total.

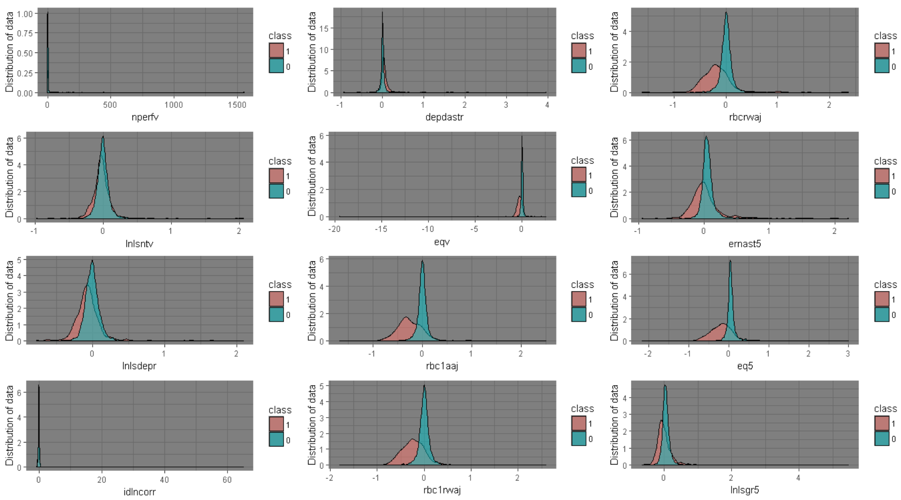


FIGURE 3.8: *Performance and Condition Ratios* group density plots: Scenario 3

The boxplots and density plots of variables from the "Performance and Condition Ratios" group, broken down by class value are presented in Figure 3.7 and

Figure 3.8 respectively. Remaing boxplots and density plots are presented in the Appendix H and Appendix I.

3.3 Dimensionality reduction and feature extraction

The paragraphs below describe the steps and calculations performed to apply the two dimensionality reduction and feature extraction methods described in the previous chapter: PCA and EFA.

3.3.1 Principal Components Analysis

PCA was conducted on normalized data. The correct number of components in PCA was identified based on the variance threshold. The number of first components which overall explain about 90% of variance in data was retained for further analysis. PCA results are determined by the dimension of the data set under study. In this case, the data sets consist of 107 exploratory variables, so, 107 PCs were originally computed.

Scenario 1: Quarterly changes in financial results

In this study, as mentioned, the threshold used was set to 90%. Even by keeping the 42 PCs, totalling a 90.12% of variability in the data, we achieved a good dimensionality reduction by explaining the variability of the 107 variables with just 42 in the new space.

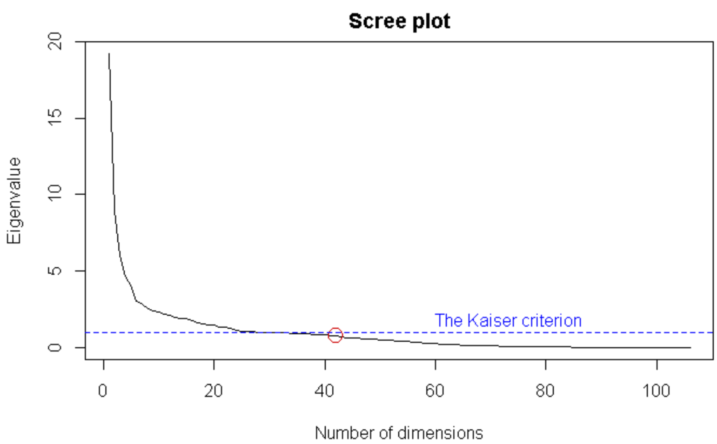


FIGURE 3.9: Scree plot: Scenario 1

The scree plot of the eigenvalues found after performing PCA on our data set is presented in the Figure 3.9. The total number of dimensions retained doesn't satisfy the Kaiser criterion. However, the lowest eigenvalue is equal to 0.76, which is still quite sufficient.

The first 3 PCs explain about 34.25% of the total variance. Therefore, I plotted each individual in these three dimensions as shown in Figure 3.10. The individuals are grouped by class, the blue triangles represent the non-bankrupt entities, while yellow circles represent the bankrupt banks. It seems that some bankrupt banks have high correlation with the third dimension while the non-bankrupt banks don't have or have very low correlation with the third dimension. However, non-bankrupt banks seem to have mostly positive correlation with the first PC, while, bankrupt banks have negative correlation with the first PC. Also, We see that there is no clear-cut linear separation between the bankrupt and non-bankrupt banks in the first three dimensions.

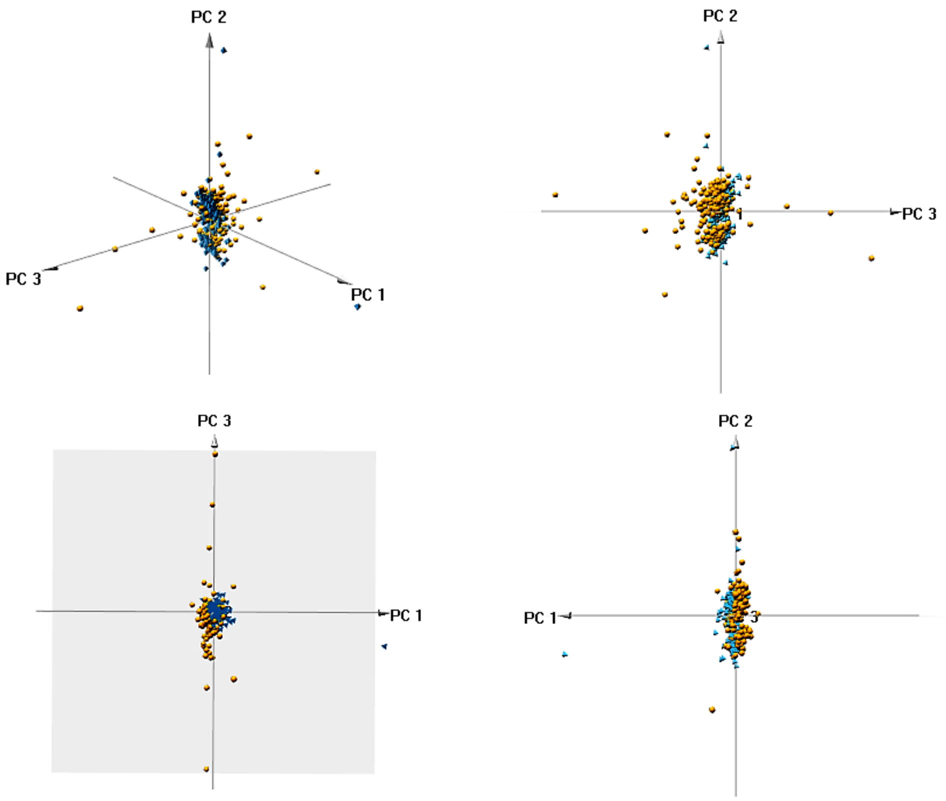


FIGURE 3.10: The first 3 dimensions obtained from PCA: Scenario 1.

3.3.2 Scenario 2: Half-year changes in financial results

The scree plot of the PCA of the data set composed from the half-year changes in financial results is presented in the Figure 3.11. The red circle on the plot denotes the number of components retained for further analysis.

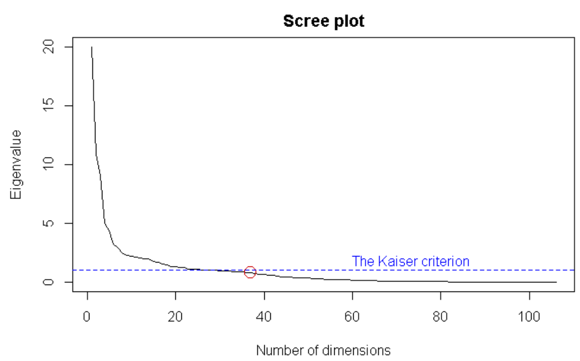


FIGURE 3.11: Scree plot: Scenario 2

Based on the threashold of 90% of variance in the data, 37 first PCs were selected. Retained dimensions in total gave us 90.7% of cumulative variance of the initial data set. The Kaiser criterion is not achieved, the last eigenvalue is equal to 0.76.

The first 3 dimensions of PCA cover about 37.73% of the total variance in the data. The individual plots represented in the first 3 dimensions are presented in Figure 3.12.

The individuals are grouped by class. The bankrupt entities' data is grouped near the center of the first three principal components, which means low correlation with the dimensions, while, the non-bankrupt entities which are presented by blue triangles are more scattered across the dimensions. However, there is no clear separation between the bankrupt and non-bankrupt banks in the first three dimensions.

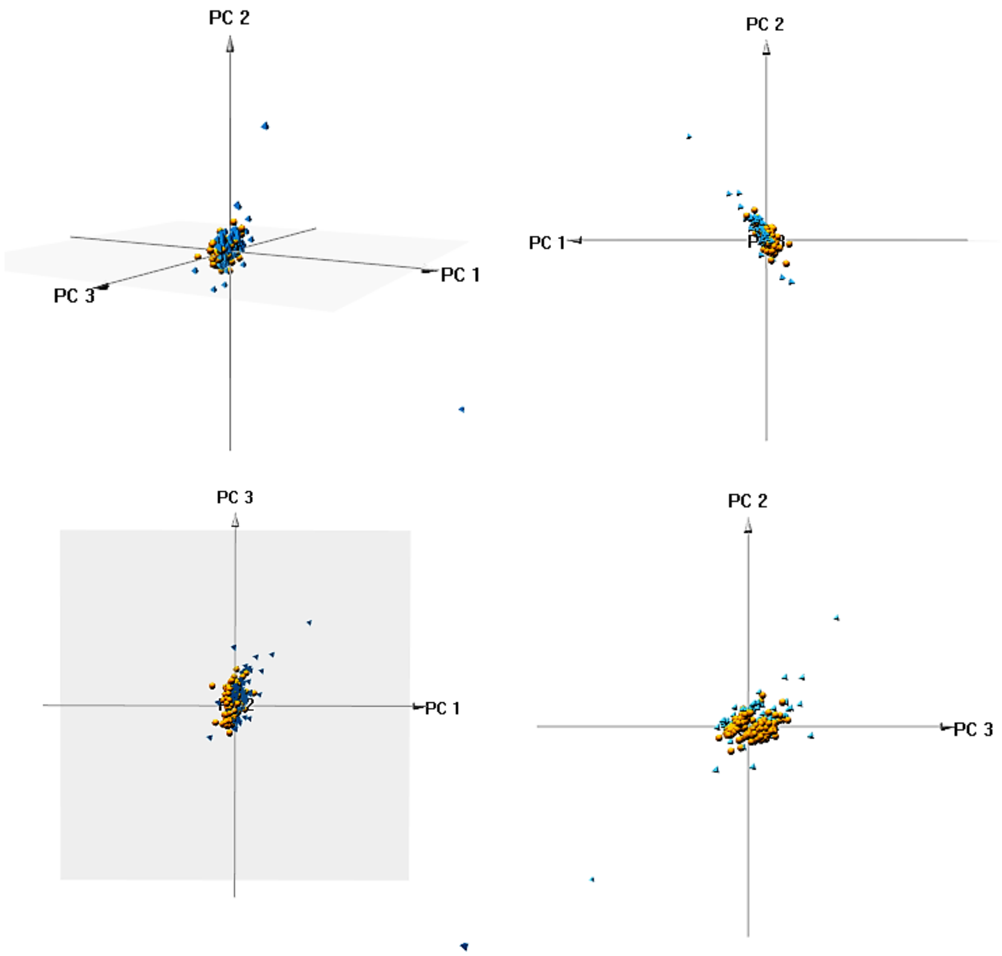


FIGURE 3.12: The first 3 dimensions obtained from PCA: Scenario 2.

3.3.3 Scenario 3: Yearly changes in financial results

The final number of dimensions retained from the PCA is equal to 36, which is almost 3 times less than initial features number. The total variance explained by 36 dimensions is equal to 90.22%.

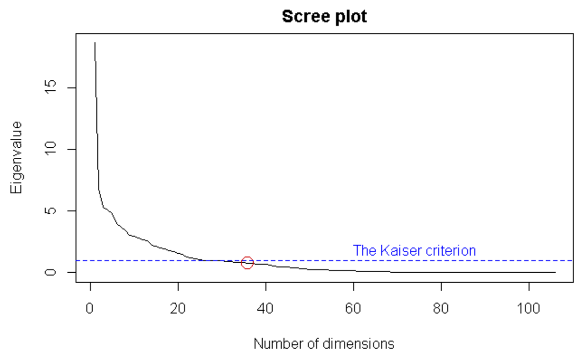


FIGURE 3.13: Scree plot: Scenario 3

The Kaiser criterion is not satisfied, as the lowest eigenvalue has a value of 0.8. The scree plot is presented in Figure 3.13.

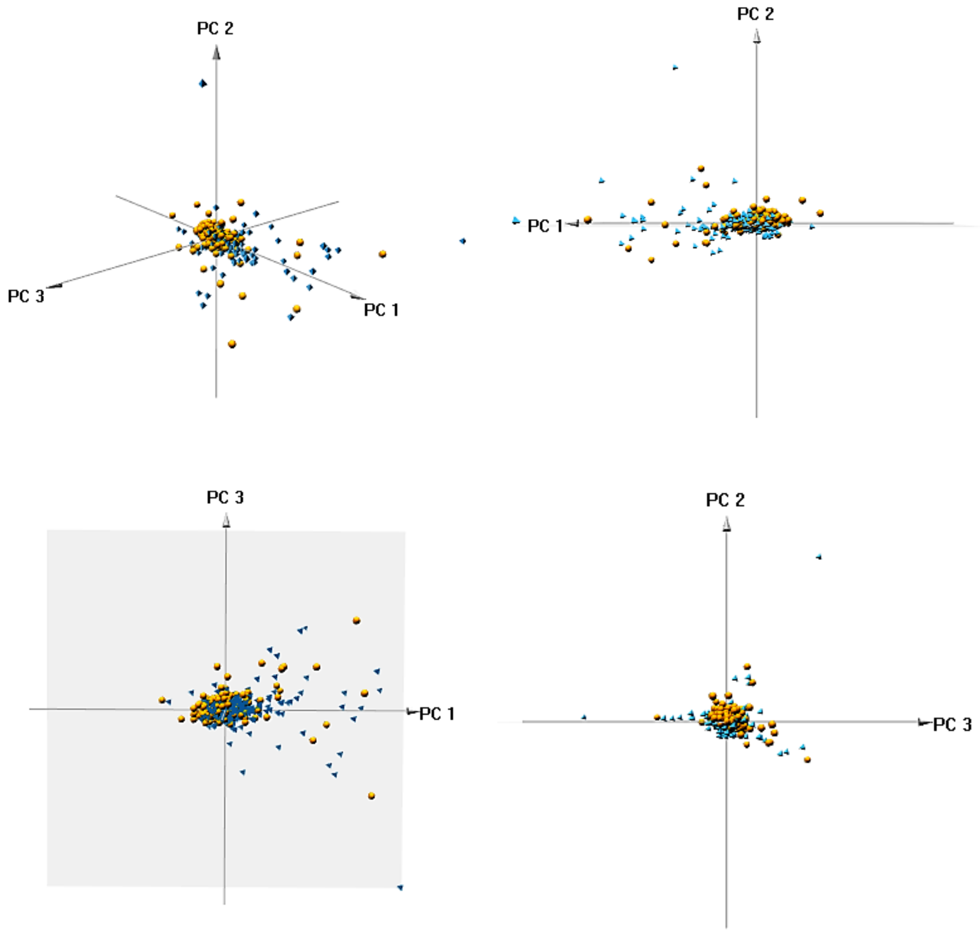


FIGURE 3.14: The first 3 dimensions obtained from PCA: Scenario 3.

The individual plots grouped by class of the first 3 dimensions of PCA are presented in Figure 3.14. The first 3 dimensions explain about 29.06% of the total variance. We can see in Figure 3.12 that there is no clear linear separation between the bankrupt and non-bankrupt banks in the first three dimensions.

3.3.4 Exploratory Factor Analysis

EFA can be conducted on data which satisfy several assumptions. Firstly, the adequacy of the sample size needs to be assessed. Overall, the sample size should be greater than 300. The Kaiser-Meyer-Olkin (KMO) test tells us whether or not enough items are predicted by each factor. A minimum acceptable score for this test is 0.5. Also, the Bartlett test should be significant, which means that the variables are highly enough correlated to provide a reasonable basis for factor analysis. The Very Simple Structure (VSS) algorithm was used to identify the optimal number of interpretable factors. The final EFA was conducted by using unrotated factor analysis, namely the "Promax" method was selected as it allows correlations between factors.

Scenario 1: Quarterly changes in Financial Results

The data set which is comprised of the quarterly changes in the financial data satisfies the main assumptions of the EFA. The KMO test which is Measure of Sampling Adequacy (MSO) is equal to 0.82, the Bartlett test is significant (p=2.22e-16).

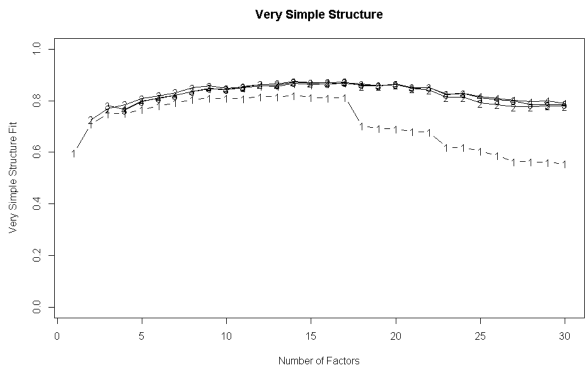


FIGURE 3.15: Very Simple Structure Fit: Scenario 1

The Very Simple Structure (VSS) results are presented in Figure 3.15. According to the plot, VSS achieves a maximum fit of 0.87 with 14 factors at complexity 2. The number of factors which maximizes the fit and explains as much variability as possible was selected. In this case, it is 17 factors, which provide on average a fit of 0.86. The cumulative variance explained by the 17 factors is equal to 61.2%. The Kaiser criterion is satisfied as the lowest eigenvalue has a value of 1.58. The Cattell scree test is not useful for this data set, since the number of dimensions to retain should be less than 10, which explains only up to 48.8% of variance in the data. The loadings of factors are presented in the Figure 3.16 and Figure 3.17. There is a pattern in the exploratory variables that have most influenced (higher loadings) the formation of factors, which leads to following interpretation of factors:

- Factor 1: Represents Assets and Liabilities, Deposits and Loans and Leases related variables.
- Factor 2: Income and Expense group variables.
- Factor 3: Loans and Leases has high positive influence, and Deposits has negative impact.
- Factor 4: Equity related financial data.

- Factor 5: Income related data have positive impact while Liabilities negatively affected.
- Factor 6: Represents Operating income results.
- Factor 7: Equity results.
- Factor 8: Noncurrent Loans and Leases.
- Factor 9: Data related to Unused Commitments.
- Factor 10: Total assets related data.
- Factor 11: Loss allowance related data.
- Factor 12: Interest bearing deposits.
- Factor 13: U.S. Government securities.
- Factor 14: Total securities.
- Factor 15: Interest income.
- Factor 16: Non-interest expense.
- Factor 17: Earning assets.

	Name	Definition	Group	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16	F17
1	eothnint	All other noninterest expense	Additional Noninterest Expense	-,06	,73	-,01	-,03	-,02	,01	-,02	-,01	,04	,02	,07	-,01	,00	-,01	,01	,11	,02
2	asset	Total assets	Assets and Liabilities	,98	,00	-,29	,00	,02	-,02	,01	,00	-,01	-,30	-,06	-,01	,00	,00	,00	-,01	,00
3	liabeq	Total liabilities and capital	Assets and Liabilities	,98	,00	-,29	,00	,02	-,02	,01	,00	-,01	-,30	-,06	-,01	,00	,00	,00	-,01	,00
4	asset2	Average assets, quarterly	Assets and Liabilities	,89	-,01	-,07	,01	,01	-,01	,02	,00	,03	-,04	-,02	-,02	-,01	,01	,00	,00	,00
5	lnlsnet	Net loans and leases	Assets and Liabilities	,87	,00	,53	-,01	,00	-,01	,00	,00	-,03	,01	-,06	-,01	,00	,00	-,01	-,01	,00
6	AVASSETJ	Adjusted average assets for leverage capital purposes	Assets and Liabilities	,86	,01	-,12	-,01	-,01	,01	,03	,01	,01	-,10	,01	-,02	,02	,00	-,01	,00	,00
7	dep	Total deposits	Assets and Liabilities	,86	,01	-,34	-,01	,02	-,01	,00	,00	-,02	,18	-,03	-,01	,00	,00	,00	-,01	,00
8	depdom	Deposits held in domestic offices	Assets and Liabilities	,85	,01	-,34	-,01	,02	-,01	,00	,00	-,02	,20	-,04	-,01	,00	,00	,00	-,01	,00
9	RWAJT	Total risk weighted assets adjusted	Assets and Liabilities	,81	,01	,25	-,04	,01	-,02	,04	-,01	,04	-,05	-,07	,04	-,01	,00	,00	-,02	,01
10	numemp	Total employees (full-time equivalent)	Assets and Liabilities	,68	,01	,02	-,03	,02	,03	,02	-,02	,17	-,03	,02	,00	,00	,02	,02	,05	,02
11	asset5	Average total assets	Assets and Liabilities	,63	-,03	,09	,05	-,01	,00	-,01	,00	,03	,35	,00	-,01	-,02	-,01	,01	-,01	,02
12	bkprem	Bank premises and fixed assets	Assets and Liabilities	,53	,00	,03	-,01	,00	,01	,00	,04	,32	-,01	,04	,00	,08	,00	-,06	,03	,00
13	eqsur	Surplus	Assets and Liabilities	,47	,05	-,02	,00	,00	,02	,00	-,02	,07	-,01	,08	,01	-,02	,02	,01	,01	-,09
14	ernast	Earning assets	Assets and Liabilities	,44	-,02	-,07	,02	,00	-,01	,00	,00	-,01	-,10	-,03	-,01	,00	,01	,01	,00	-,01
15	liab	Total Liabilities	Assets and Liabilities	,36	,02	-,12	-,03	-,41	,02	-,01	-,01	-,02	-,10	-,06	,00	,00	,00	,00	,00	,01
16	oaienc	Income earned, not collected on loans	Assets and Liabilities	,30	,05	,16	,04	,00	,00	-,01	,01	-,01	-,04	-,03	,03	,00	,01	,03	-,05	,07
17	idoa	All other assets	Assets and Liabilities	,23	-,01	,12	-,01	,01	,02	,02	-,02	,01	-,04	,02	,01	,00	-,01	,03	-,01	,02
18	idoliab	All other liabilities	Assets and Liabilities	,12	,03	-,01	,01	-,17	,01	,00	,01	-,01	-,05	-,02	,01	,00	,00	-,01	,05	,01
19	RBCT2	Tier 2 Risk-based capital	Assets and Liabilities	,10	-,06	,02	,07	,00	-,01	,02	,00	,00	-,02	,01	,02	,00	,01	-,02	,03	,25
20	lnatres	Loan loss allowance	Assets and Liabilities	,08	-,02	,02	-,01	,00	-,01	,00	-,01	-,02	-,02	,97	,01	,00	,00	,00	-,02	,00
21	voliab	Volatile liabilities	Assets and Liabilities	,08	,07	,02	-,01	-,01	,00	,00	,01	-,01	-,06	-,03	,00	,00	-,01	,00	,00	,01
22	eq	Bank equity capital	Assets and Liabilities	,06	,01	,00	,06	-,01	,00	,96	,00	,00	,00	,00	,00	,00	,00	,00	,00	-,01
23	eqtot	Total equity capital	Assets and Liabilities	,06	,01	,00	,06	-,01	,00	,96	,00	,00	,00	,00	,00	,00	,00	,00	,00	-,01
24	chbal	Cash & Balances due from depository institutions	Assets and Liabilities	,03	,01	-,16	-,01	,01	,00	,00	-,02	-,01	-,07	,07	-,01	,00	-,02	,00	,00	,00
25	nclnls	Noncurrent loans and leases	Assets and Liabilities	,02	,00	,00	,00	,00	,00	,00	,99	,00	,00	,01	,00	,00	,00	,00	,00	,00
26	equptot	Undivided profits	Assets and Liabilities	,01	-,02	-,01	-,01	,00	,09	,01	,01	,00	,00	-,03	,00	,00	,01	,00	,03	,13
27	RBCT1J	Tier one (core) capital	Assets and Liabilities	,00	-,01	,01	,98	,00	,00	,01	,00	,01	-,01	,00	,00	,00	,00	,00	,00	,00
28	sc	Total securities	Assets and Liabilities	-,01	,00	-,01	,00	,00	,00	,00	,00	,00	,00	,00	,00	,00	,99	,00	,00	,00
29	depi	Interest-bearing deposits	Assets and Liabilities	-,05	,00	,00	,00	,00	,00	,00	,00	,00	-,01	,00	,99	,00	,00	,00	,00	,00
30	uc	Total unused commitments	Assets and Liabilities	-,07	-,02	-,03	,00	,00	-,01	,00	,00	1,00	,01	-,01	,00	-,01	,00	,01	-,01	,00
31	ucln	Unused loan commitments	Assets and Liabilities	-,07	-,02	-,03	,00	,00	-,01	,00	,00	1,00	,01	-,02	,00	-,01	,00	,01	-,01	,00
32	LNRERSFM	Loans secured by 1-4 family first liens	1-4 Family Residential Net Loans and Leases	,45	,02	,12	,01	,02	,01	-,02	,00	,00	-,01	,05	,01	,00	,01	,02	,02	-,01
33	chbalni	Total noninterest-bearing balances	Cash and Balances Due	,01	,05	-,13	-,01	,02	-,01	-,02	-,01	,05	,00	,01	,00	,03	-,03	-,01	-,02	,00
34	eqcprev	Amended balance at previous year-end	Changes in Bank Equity Capital	,10	,11	,02	,16	-,05	,01	-,16	,01	,02	-,02	,00	,00	,00	-,01	,01	-,03	,01
35	netinc	Net income	Changes in Bank Equity Capital	,02	,05	,01	,02	,78	,02	,01	,02	-,01	-,03	,05	,00	,00	,00	,00	,00	-,01
36	idothnii	Additional Noninterest Income	Income and Expense	,04	-,01	,04	,02	,00	,01	,01	,00	-,02	-,01	,05	,00	,00	,00	,03	,00	,00
37	esal	Salaries and employee benefits	Income and Expense	,03	1,00	,02	-,01	,00	,00	,00	,00	,01	,00	-,02	,00	,00	,00	-,01	-,03	-,06
38	ibefxtr	Income before extraordinary items	Income and Expense	,03	,07	,01	,01	,58	,04	,00	,01	-,01	-,04	,00	,00	,00	,00	,00	-,01	,00
39	nonii	Total noninterest income	Income and Expense	,02	,24	,07	,03	-,02	,00	,01	,01	-,07	-,02	,00	,00	-,01	,01	,00	-,01	-,02
40	nonix	Total noninterest expense	Income and Expense	-,01	,51	,00	,03	,01	-,01	-,01	,00	-,02	-,01	-,03	,00	,00	,00	,00	,72	-,12
41	eintexp	Total interest expense	Income and Expense	-,02	1,01	,00	,00	,01	,00	,01	,00	-,01	,00	,00	,00	,00	,00	-,01	-,05	,01
42	idpretx	Pre-tax net operating income	Income and Expense	-,02	,00	-,01	,00	-,02	,99	,00	,00	,00	,00	,00	,00	,00	,00	,00	-,01	-,01
43	intinc	Total interest income	Income and Expense	-,02	1,01	,01	,00	,00	-,01	,01	,00	-,01	,00	-,01	,00	,00	,00	,00	-,06	,02
44	nim	Net interest income	Income and Expense	-,02	,24	-,01	,00	,00	,00	,01	,00	,00	,00	,00	,00	,00	,00	,92	-,02	-,07
45	noij	Net operating income	Income and Expense	-,02	,00	-,01	,00	-,01	,99	,00	,00	,00	,00	,00	,00	,00	,00	,00	-,01	-,01
46	epremagg	Premises and equipment expense	Income and Expense	-,03	,94	,00	,00	,00	,00	,00	,00	-,01	,00	,01	,00	,01	,00	-,01	-,04	,02
47	IDEOTH	Additional noninterest expense	Income and Expense	-,06	,60	,01	-,03	-,02	,01	-,02	-,01	,03	,02	,05	-,01	,00	-,01	,01	,11	-,01
48	lnls	Loans and leases, gross	Maturity & Repricing for Loans and Leases	,87	,00	,52	-,02	,00	-,01	,00	,00	-,03	,01	,00	-,01	,00	,00	,00	-,01	-,01
49	idlnls	Loans and leases, gross	Net Loans and Leases	,87	,00	,52	-,02	,00	-,01	,00	,00	-,03	,01	,00	-,01	,00	,00	,00	-,01	-,01
50	lnlsgr	Total loans and leases	Net Loans and Leases	,87	,00	,52	-,02	,00	-,01	,00	,00	-,03	,01	,00	-,01	,00	,00	,00	-,01	-,01
51	lnre	All real estate loans	Net Loans and Leases	,74	-,03	,35	,03	,00	,01	-,02	,00	,00	-,01	,05	,00	,00	,00	-,01	-,01	-,02
52	lnredom	Real estate loans in domestic offices	Net Loans and Leases	,74	-,03	,34	,03	,00	,01	-,02	,00	,00	-,01	,05	,00	,00	,00	-,01	-,01	-,02
53	lnreres	1-4 family residential loans	Net Loans and Leases	,68	,01	,19	,03	,04	,01	-,04	,00	,01	-,01	,07	,00	,00	,01	,02	,04	-,01

FIGURE 3.16: Exploratory Factor Analysis: Scenario 1 (1 Part)

	Name	Definition	Group	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16	F17
54	lnconoth	Other loans to individuals	Net Loans and Leases	,47	,09	,03	-,03	,00	,01	,00	,00	-,01	-,03	,01	,01	,00	,01	,00	,02	,03
55	lnci	Commercial and industrial loans	Net Loans and Leases	,15	-,02	,15	-,01	-,02	-,01	-,01	,00	-,02	-,04	,02	-,01	,00	,00	,00	,00	,01
56	lnrenres	Secured by nonfarm nonresidential properties	Net Loans and Leases	,07	,02	,00	-,01	,00	,00	-,01	,00	-,01	,00	-,04	,00	,00	,00	-,01	,00	-,02
57	lncon	Loans to individuals	Net Loans and Leases	-,01	,02	-,05	-,02	,02	,00	-,01	-,01	,11	,00	,02	,00	,00	-,01	,00	-,01	,01
58	ntripc	Individuals, partnerships, and corporations	Nontransaction Accounts	,82	-,04	-,16	-,02	,00	,01	,02	,00	-,04	,03	,00	,00	,00	,00	,01	,01	,02
59	ernast5	Average earning assets	Performance and Condition Ratios	,77	-,01	-,01	,05	-,01	,01	-,02	,00	,00	-,07	,04	,00	-,02	-,01	,00	,00	,00
60	lnlsgr5	Average total loans	Performance and Condition Ratios	,74	-,01	,16	,02	-,01	,02	-,03	,00	-,03	-,10	,04	,00	-,02	-,03	,00	,00	,05
61	eq5	Average equity	Performance and Condition Ratios	,28	,11	,03	,22	-,03	-,02	-,08	,01	,06	-,08	-,07	,01	-,01	,01	,01	-,07	,01
62	astempm	Assets per employee ($millions)	Performance and Condition Ratios	,10	-,02	-,19	,03	-,01	-,04	-,04	,03	-,09	-,06	-,03	-,01	-,02	-,02	-,02	-,05	,00
63	roeinjr	Retained earnings to average equity (ytd only)	Performance and Condition Ratios	,06	-,05	-,04	-,01	,76	-,01	-,02	-,03	,00	,00	-,10	,01	,00	,00	,00	,01	,01
64	roe	Return on Equity (ROE)	Performance and Condition Ratios	,05	-,01	-,03	-,01	,97	-,02	-,02	-,01	,00	,00	-,03	,00	,00	,00	,00	,00	,00
65	roa	Return on assets (ROA)	Performance and Condition Ratios	,04	,01	,00	,00	,93	-,01	,01	,01	,00	-,01	,03	,00	,00	,00	,00	,01	,00
66	noniiay	Noninterest income to average assets	Performance and Condition Ratios	,03	,05	,03	,04	-,04	,00	,03	,01	-,04	-,01	-,02	,00	-,01	,01	,00	,00	,04
67	intincy	Yield on earning assets	Performance and Condition Ratios	,02	-,04	-,01	,00	,00	-,01	-,02	,00	,00	,01	-,03	,00	,00	,01	,03	,07	,98
68	nperfv	Noncurrent assets plus other real estate owned	Performance and Condition Ratios	,02	,01	,01	-,01	,01	-,01	,00	,89	,00	,02	-,08	,01	,00	,00	,00	,01	,01
69	roaptx	Pretax return on assets	Performance and Condition Ratios	,02	-,01	,00	,00	,14	,03	,00	,00	-,01	-,02	-,01	,00	,00	,00	,00	-,01	,02
70	idlncorr	Net loans and leases to core deposits	Performance and Condition Ratios	,01	,06	,65	-,04	,00	-,01	,02	-,01	-,01	-,17	-,04	,00	,00	-,01	,01	,00	,00
71	eqv	Equity capital to assets	Performance and Condition Ratios	,00	,01	,01	,07	-,01	,00	,95	,00	,00	,01	,00	,00	,00	,00	,00	,00	-,01
72	lnlsdepr	Net loans and leases to deposits	Performance and Condition Ratios	,00	-,01	,94	,01	,00	-,01	,00	,00	-,01	-,12	-,01	,00	,00	,00	-,01	,00	-,01
73	eeffr	Efficiency ratio	Performance and Condition Ratios	-,01	-,06	-,02	,00	,01	,00	-,02	,00	-,01	-,01	,02	,00	,00	-,02	-,01	,21	-,07
74	lnresncr	Loan loss allowance to noncurrent loans	Performance and Condition Ratios	-,01	,01	-,01	,00	,01	-,01	,00	-,05	,00	-,01	,02	,00	,00	,00	,00	-,01	,03
75	nclnlsr	Noncurrent loans to loans	Performance and Condition Ratios	-,01	,00	-,01	,01	-,02	,00	,00	,98	,00	,00	,07	,00	,00	,00	,00	,00	-,01
76	nimy	Net interest margin	Performance and Condition Ratios	-,01	-,20	-,01	,00	,00	,00	-,01	,00	,01	,00	,00	,00	,00	,00	1,00	,02	,08
77	noijy	Net operating income to assets	Performance and Condition Ratios	-,02	-,01	-,01	,00	,01	,99	,00	,00	,00	,00	-,02	,00	,00	,00	,00	,00	,00
78	nonixay	Noninterest expense to average assets	Performance and Condition Ratios	-,02	-,24	-,01	,01	,01	-,01	,01	,00	-,02	,00	-,02	,00	,00	,00	,00	1,02	,11
79	rbc1aaj	Core capital (leverage) ratio	Performance and Condition Ratios	-,05	,00	,01	,98	,00	,00	,03	,00	,00	,00	,00	,00	,00	,00	,00	,01	,00
80	intexpy	Cost of funding earning assets	Performance and Condition Ratios	-,06	,24	-,04	-,02	,01	,01	,02	-,01	,01	,03	,06	,00	,00	-,01	-,02	,03	,68
81	rbc1rwaj	Tier 1 risk-based capital ratio	Performance and Condition Ratios	-,06	-,01	-,01	1,00	,00	,00	,00	,00	-,01	,00	,00	,00	,00	,00	,00	,01	,00
82	rbcrwaj	Total risk-based capital ratio	Performance and Condition Ratios	-,06	-,02	-,01	,96	,01	,00	,02	,00	,00	,00	,01	,00	,00	,00	-,01	,01	,00
83	lnlsntv	Net loans and leases to total assets	Performance and Condition Ratios	-,12	,00	,98	,00	-,02	,00	,00	,00	-,01	,63	-,04	,00	,00	,00	,00	,00	,00
84	depdastr	Total domestic deposits to total assets	Performance and Condition Ratios	-,14	,01	,06	,00	-,02	,00	,01	,00	,01	1,02	-,02	-,01	,00	,00	,00	,00	,01
85	lnatresr	Loss allowance to loans	Performance and Condition Ratios	-,24	-,02	-,13	,01	-,02	-,01	,00	,03	,00	-,03	,99	,00	,00	,00	,00	-,02	-,01
86	scage	U.S. Government agency obligations	Securities	,00	,00	,00	,00	,00	,00	,00	,00	-,01	,00	,00	,00	,99	,00	,00	,00	,00
87	scus	U.S. Government securities	Securities	,00	,00	,00	,00	,00	,00	,00	,00	-,01	,00	,00	,00	,99	,00	,00	,00	,00
88	scrdebt	Total debt securities	Securities	-,01	,00	-,01	,00	,00	,00	,00	,00	,00	,00	,00	,00	,00	,99	,00	,00	,00
89	ntrtmlg	Amount ($) - time deposits of $100,000	Time Deposits at the $100,000 Threshold	,00	,00	-,02	,00	,02	,00	,01	-,01	,00	,00	,07	,00	,00	-,01	,00	,01	-,01
90	ts	Total time and savings deposits	Total Deposits	,88	-,04	-,26	-,03	,02	,00	,01	,01	-,04	,04	-,01	,00	,00	,00	,00	,01	,01
91	ntr	Nontransaction accounts	Total Deposits	,86	-,04	-,23	-,04	,01	,01	,02	,02	-,04	,06	,00	,00	-,01	,00	,00	,01	,00
92	ntrtime	Total time deposits	Total Deposits	,73	-,01	-,16	-,06	,02	,00	,00	,03	-,05	,04	,01	-,01	,01	-,05	,00	,00	,01
93	idtrni	Individuals, partnerships, and corporations	Total Deposits	,67	,02	-,16	,00	-,01	,00	,01	,00	-,01	,06	-,01	-,01	,01	,00	,01	,01	-,03
94	coredep	Retail deposits	Total Deposits	,63	-,08	-,25	,03	,02	,00	-,02	,02	-,02	,15	,06	-,01	,00	,01	-,01	,00	-,02
95	depins	Estimated insured deposits	Total Deposits	,58	,02	-,14	-,04	,01	,00	,02	-,02	-,03	,09	,03	-,01	,01	,00	-,01	,00	,00
96	irakeogh	IRAs and Keogh plan accounts	Total Deposits	,52	,00	-,01	-,01	,00	,01	,00	-,01	-,06	,01	,05	,01	,00	,01	,01	,02	,01
97	ntrsoth	Other savings deposits (excluding MMDAs)	Total Deposits	,33	-,01	-,05	,00	,00	,01	,01	-,01	-,02	,00	,02	,03	-,02	,02	,00	,02	,00
98	trn	Transaction accounts	Total Deposits	,33	,07	-,28	,05	,01	-,02	-,04	-,02	,00	,22	-,02	,07	,00	,01	-,02	-,02	-,01
99	ddt	Demand deposits	Total Deposits	,23	,10	-,24	,04	,00	-,01	-,01	-,02	,00	,13	-,02	,05	,00	,00	-,01	-,01	-,02
100	ntrsmmda	Money market deposit accounts (MMDAs)	Total Deposits	,15	,02	-,09	,01	-,01	,00	,02	-,01	,00	-,01	-,04	,05	,00	,01	,00	,00	-,01
101	depnidom	Noninterest-bearing deposits	Total Deposits	,14	,08	-,18	,02	,01	-,01	,00	-,01	,00	,12	-,02	-,07	,00	,00	,05	-,02	-,02
102	depidom	Interest-bearing deposits	Total Deposits	-,05	,00	,00	,00	,00	,00	,00	,00	,00	-,01	,00	,99	,00	,00	,00	,00	,00
103	edepdom	Interest expense: Domestic office deposits	Total Interest Expense	-,01	1,01	,00	,00	,01	,00	,01	,00	-,01	,01	,00	,00	,00	,00	-,01	-,05	,00
104	ilndom	Interest income: Domestic office loans	Total Interest Income	-,02	,96	-,02	-,01	,01	-,01	,01	,00	-,01	,00	-,04	,00	,00	,00	-,01	-,06	,07
105	isc	Interest income: Securities	Total Interest Income	-,02	,97	,00	,01	,00	,00	,01	,00	-,01	,00	-,02	,00	,00	,01	,00	-,06	-,02
106	trnipcoc	Individuals, partnerships and corporations	Transaction Accounts	,12	,06	-,08	,02	,00	,00	-,02	-,01	,00	,03	-,01	,03	,01	,01	-,01	,00	-,01

FIGURE 3.17: Exploratory Factor Analysis: Scenario 1 (2 Part)

Scenario 2: Half-year changes in financial results

The sampling adequacy test of half-year financial data provided a result of 0.84, which is enough for EFA. The Bartlett test is also significant (p=2.22e-16).

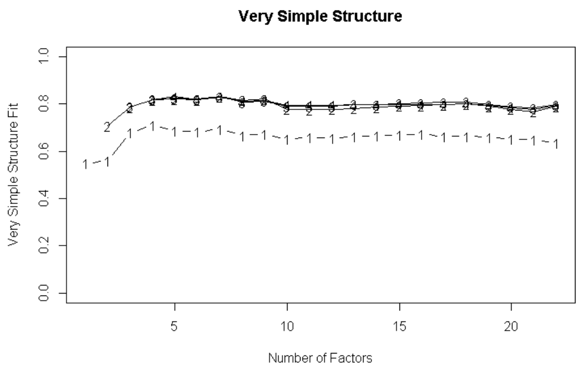


FIGURE 3.18: Very Simple Structure Fit: Scenario 2

The VSS results are presented in Figure 3.18. The plot of VSS fit does not provide the exact optimal number of factors. Based on the total variance explained by factors and VSS fit, 17 factors were selected for EFA. The retained factors explain 66.7% of variance in the data set and achieved average a fit of 0.79. The last eigenvalue in EFA is equal to 1.3 which means that Kaiser criterion is satisfied. The obtained factors' loadings are presented in the Figure 3.19 and Figure 3.20. The relation between factors obtained from EFA are exploratory variables is summarized below:

- Factor 1: Represents Assets and Liabilities, Deposits and Loans and Leases related variables.
- Factor 2: Assets and Deposits group variables.
- Factor 3: Income and Expense related variables.
- Factor 4: Equity (capital) related financial data.
- Factor 5: Income (earnings) related data.
- Factor 6: Assets and Equity data.
- Factor 7: Assets and Deposits group variables.
- Factor 8: Securities.
- Factor 9: Loans and Leases.
- Factor 10: Operating income related variables.
- Factor 11: Domestic deposits related data.
- Factor 12: Loss allowance to loans and Non-current assets and other real estate.
- Factor 13: Non-current loans data.
- Factor 14: Unused commitments.

- Factor 15: Non-interest income.
- Factor 16: Interest income.
- Factor 17: Interest-bearing deposits.

Name	Definition	Group	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16	F17
1 eothnint	All other noninterest expense	Additional Noninterest Expense	-,03	,00	,95	-,10	,00	,00	,01	,01	-,01	-,01	,02	,00	,00	,00	,00	-,01	,00
2 asset	Total assets	Assets and Liabilities	,32	,72	-,03	-,01	-,01	-,03	,03	,00	-,03	,00	-,22	-,03	,00	,00	,00	,01	-,01
3 asset2	Average assets, quarterly	Assets and Liabilities	-,18	,71	,01	-,01	,01	,34	,00	,00	,01	,00	-,20	-,01	,00	-,01	,00	-,03	-,01
4 asset5	Average total assets	Assets and Liabilities	-,11	,22	,00	-,06	,01	,90	,00	,01	-,03	,00	-,02	,00	,00	,00	,00	,00	,00
5 AVASSETJ	Adjusted average assets for leverage capital purposes	Assets and Liabilities	,31	,63	,00	-,01	,00	,14	,04	,01	-,03	,00	-,13	,01	,01	,00	,00	-,02	-,02
6 bkprem	Bank premises and fixed assets	Assets and Liabilities	-,07	,23	-,02	-,01	-,01	-,01	,02	,00	,16	,01	-,03	,03	-,01	,01	,00	,00	-,01
7 chbal	Cash & Balances due from depository institutions	Assets and Liabilities	-,06	,12	-,01	-,07	,00	-,05	,32	-,02	-,09	,00	-,12	,06	-,01	-,03	,01	-,03	,00
8 dep	Total deposits	Assets and Liabilities	-,07	1,09	,00	,02	,00	-,01	,05	,00	-,02	,00	,29	-,02	,00	,01	,00	-,01	-,01
9 depdom	Deposits held in domestic offices	Assets and Liabilities	-,07	1,10	,00	,02	,00	-,01	,05	,00	-,02	,00	,31	-,02	,00	,01	,00	-,01	-,01
10 depi	Interest-bearing deposits	Assets and Liabilities	,02	,44	,00	,00	,00	,00	,08	,00	-,01	,00	-,02	,00	,00	,00	,00	,00	,77
11 eq	Bank equity capital	Assets and Liabilities	,63	,10	-,01	,45	,01	-,03	,00	,00	-,04	-,01	-,04	,02	,00	,00	,00	,01	,00
12 eqsur	Surplus	Assets and Liabilities	,09	,00	-,01	,00	,00	,00	,00	,00	,02	,00	-,01	,01	,00	,01	,00	-,01	,00
13 eqtot	Total equity capital	Assets and Liabilities	,63	,10	-,01	,45	,01	-,03	,00	,00	-,04	-,01	-,04	,02	,00	,00	,00	,01	,00
14 equptot	Undivided profits	Assets and Liabilities	-,03	,00	,02	,09	-,01	,01	-,01	,00	,01	,00	,02	,01	,00	,00	,01	,00	,00
15 ernast	Earning assets	Assets and Liabilities	,19	,29	-,01	-,03	,00	,31	,00	,00	-,03	,00	-,11	-,03	,00	,00	,00	-,04	-,01
16 idoa	All other assets	Assets and Liabilities	,07	,15	,10	,08	,02	-,05	,04	-,02	,07	,00	-,16	,06	-,02	,00	,00	-,01	,00
17 idoliab	All other liabilities	Assets and Liabilities	-,06	,13	,03	-,05	-,01	-,03	,00	-,01	-,04	,01	-,08	-,02	-,01	,01	,00	-,01	,00
18 liab	Total Liabilities	Assets and Liabilities	,02	,91	-,02	-,14	-,01	-,03	,03	,00	,02	,00	-,21	-,03	,00	,00	-,01	,00	-,02
19 liabeq	Total liabilities and capital	Assets and Liabilities	,32	,72	-,03	-,01	-,01	-,03	,03	,00	-,03	,00	-,22	-,03	,00	,00	,00	,01	-,01
20 lnatres	Loan loss allowance	Assets and Liabilities	,01	,00	,02	-,03	,00	,00	-,04	,00	,04	,01	,02	,98	-,04	,00	,00	-,01	,00
21 lnlsnet	Net loans and leases	Assets and Liabilities	1,03	-,02	-,01	-,04	,00	-,02	,01	,00	,03	,00	,02	-,01	,00	,00	,00	,00	,00
22 nclnls	Noncurrent loans and leases	Assets and Liabilities	,00	,01	,01	-,01	-,01	,00	,01	,00	,00	-,01	,00	-,04	,99	,00	,00	,00	,00
23 numemp	Total employees (full-time equivalent)	Assets and Liabilities	-,06	,23	,00	,06	-,01	,04	,48	-,01	,09	,00	-,07	,05	,00	,04	,01	,04	,02
24 oaienc	Income earned, not collected on loans	Assets and Liabilities	,89	,03	,09	-,02	,00	-,03	,02	,00	,05	,00	-,01	,00	-,01	-,02	,00	,00	,00
25 RBCT1J	Tier one (core) capital	Assets and Liabilities	,66	,05	-,01	,46	,00	-,03	,00	,00	-,04	,01	-,03	-,02	,00	,01	,00	,02	,00
26 RBCT2	Tier 2 Risk-based capital	Assets and Liabilities	1,05	-,11	-,03	-,03	,00	-,03	,02	,00	-,10	,00	,03	,03	,00	,00	,00	,03	,00
27 RWAJT	Total risk weighted assets adjusted	Assets and Liabilities	,38	,46	,01	-,08	,02	-,01	-,05	-,03	,18	,00	-,18	,01	,00	,01	,00	,06	-,02
28 sc	Total securities	Assets and Liabilities	,01	-,03	,00	,02	,00	,02	,00	,99	,01	,00	,02	,01	,00	,01	,00	,00	,00
29 uc	Total unused commitments	Assets and Liabilities	,00	,01	-,01	,00	,00	,00	,00	,01	,00	,00	,01	,00	,00	,99	,00	,00	,00
30 ucln	Unused loan commitments	Assets and Liabilities	,00	,01	,00	,00	,00	,00	,00	,01	,00	,00	,01	,00	,00	,99	,00	,00	,00
31 voliab	Volatile liabilities	Assets and Liabilities	,90	-,07	,01	-,07	-,01	-,04	,00	,00	-,09	,00	-,07	,01	-,01	,02	,00	,02	,01
32 LNRERSFM	Loans secured by 1-4 family first liens	1-4 Family Residential Net Loans and Leases	-,07	,09	,03	-,01	-,01	-,01	-,03	-,02	,28	,01	-,06	,01	,00	,00	,00	,01	,01
33 chbalni	Total noninterest-bearing balances	Cash and Balances Due	-,05	,04	,01	-,07	,01	-,04	,01	-,02	-,05	,00	-,17	,04	-,01	-,01	,01	-,04	,01
34 eqcprev	Amended balance at previous year-end	Changes in Bank Equity Capital	,01	-,27	,08	-,05	,01	,79	,03	,01	-,05	,00	,11	-,01	,00	-,01	,00	-,01	,02
35 netinc	Net income	Changes in Bank Equity Capital	-,01	-,01	,02	,02	,96	,01	,00	,00	-,01	,08	,01	,00	-,01	,00	,00	,00	-,01
36 eintexp	Total interest expense	Income and Expense	-,04	,00	1,00	,02	,00	,01	-,01	,00	,00	,00	-,01	,00	,00	,00	,00	-,03	,00
37 epremagg	Premises and equipment expense	Income and Expense	,02	-,01	,96	,02	,01	,01	-,01	-,01	-,02	,00	-,01	,01	,00	,00	,00	-,02	,00
38 esal	Salaries and employee benefits	Income and Expense	-,06	-,05	,98	,02	,00	,02	,02	-,01	-,02	,01	-,05	,01	,00	,00	,00	-,01	,00
39 ibefxtr	Income before extraordinary items	Income and Expense	-,01	-,01	,02	,02	,96	,01	,00	,00	-,01	,08	,01	,00	-,01	,00	,00	,00	-,01
40 IDEOTH	Additional noninterest expense	Income and Expense	-,01	,02	,86	-,08	-,01	,01	,02	,00	-,02	-,01	,06	-,01	,00	,00	,00	-,02	,00
41 idothnii	Additional Noninterest Income	Income and Expense	-,05	-,05	,16	,00	-,03	,00	,03	-,08	,00	-,01	-,12	,00	,00	,01	,05	-,01	,01
42 idpretx	Pre-tax net operating income	Income and Expense	,00	,00	,01	,00	-,06	,00	,00	,00	,00	,98	,00	,02	-,04	,00	,00	,00	-,01
43 intinc	Total interest income	Income and Expense	-,02	,00	1,00	,04	,01	,01	-,01	,00	-,01	,00	,00	,00	,00	,00	,00	-,03	,00
44 nim	Net interest income	Income and Expense	,01	,01	,56	-,01	,01	-,04	-,02	,01	-,01	,00	-,01	,00	,00	-,01	,00	,66	-,01
45 noij	Net operating income	Income and Expense	,00	,00	,00	,00	,12	,00	,00	,00	-,01	,93	,01	,00	-,02	,00	,00	,00	,00
46 nonii	Total noninterest income	Income and Expense	,00	-,01	,02	,01	,00	,00	-,01	-,01	,00	,00	-,01	,00	,00	,00	,98	,00	,00
47 nonix	Total noninterest expense	Income and Expense	-,06	-,01	,98	-,03	,00	,01	,01	,00	-,02	,00	,00	,00	,00	,00	,00	-,02	,00
48 lnls	Loans and leases, gross	Maturity & Repricing for Loans and Leases	1,03	-,01	-,01	-,04	,00	-,02	,01	,00	,03	,00	,02	,00	,00	,00	,00	,00	,00
49 idlnls	Loans and leases, gross	Net Loans and Leases	1,03	-,01	-,01	-,04	,00	-,02	,01	,00	,03	,00	,02	,00	,00	,00	,00	,00	,00
50 lnci	Commercial and industrial loans	Net Loans and Leases	1,06	-,13	-,03	-,05	,00	-,04	,02	,01	-,09	,00	,02	,01	,00	,00	,00	,03	,00
51 lncon	Loans to individuals	Net Loans and Leases	-,10	,16	,08	,05	,01	-,01	,00	,00	,05	,00	-,01	-,02	,00	,00	,01	-,04	,00
52 lnconoth	Other loans to individuals	Net Loans and Leases	-,03	,15	,12	,02	,01	,06	,01	,00	,03	,01	,01	-,01	-,02	,01	,01	,01	,00
53 lnlsgr	Total loans and leases	Net Loans and Leases	1,03	-,01	-,01	-,04	,00	-,02	,01	,00	,03	,00	,02	,00	,00	,00	,00	,00	,00

FIGURE 3.19: Exploratory Factor Analysis: Scenario 2 (1 Part)

	Name	Definition	Group	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16	F17
54	lnre	All real estate loans	Net Loans and Leases	-,03	,07	-,05	,00	-,01	-,03	,02	,01	,96	,00	-,04	,00	,00	,00	,00	,00	,00
55	lnredom	Real estate loans in domestic offices	Net Loans and Leases	-,03	,07	-,05	,00	-,01	-,03	,02	,01	,96	,00	-,04	,00	,00	,00	,00	,00	,00
56	lnrenres	Secured by nonfarm nonresidential properties	Net Loans and Leases	,02	-,03	-,02	-,01	,00	,00	,01	,01	,32	,00	-,03	,01	,00	-,01	,00	,01	-,01
57	lnreres	1-4 family residential loans	Net Loans and Leases	-,06	,06	,03	,02	-,01	,01	-,02	-,01	,33	,02	-,02	-,01	,00	-,01	,00	,02	,01
58	ntripc	Individuals, partnerships, and corporations	Nontransaction Accounts	-,09	,88	-,01	,01	,00	,00	-,16	-,01	,10	,01	,09	,00	,00	,00	,01	,01	,02
59	astempm	Assets per employee ($millions)	Performance and Condition Ratios	,31	,27	,02	-,02	,00	-,06	-,13	,03	-,14	,01	-,04	-,03	,01	-,04	-,01	-,12	-,07
60	depdastr	Total domestic deposits to total assets	Performance and Condition Ratios	,07	,58	,02	,01	,01	,00	,07	,02	-,05	,00	1,11	,04	,00	,01	,01	,02	,00
61	eeffr	Efficiency ratio	Performance and Condition Ratios	,03	,02	-,04	-,12	,00	,00	,00	-,01	-,03	,00	-,01	,03	,00	,00	,01	,06	,00
62	eq5	Average equity	Performance and Condition Ratios	-,04	,11	,08	,39	-,01	,46	,01	-,01	,04	,00	-,13	,05	,00	,02	-,01	,01	,02
63	eqv	Equity capital to assets	Performance and Condition Ratios	-,02	-,06	,02	,97	,02	-,02	-,03	,00	,04	-,03	,00	,02	,00	-,01	,00	-,01	-,01
64	ernast5	Average earning assets	Performance and Condition Ratios	-,10	,16	,01	-,05	,01	,94	-,02	,02	-,02	,00	-,01	-,01	,00	,00	,00	,01	,00
65	idlncorr	Net loans and leases to core deposits	Performance and Condition Ratios	-,02	-,03	-,01	-,09	,00	-,12	,00	-,01	,03	,00	-,12	,00	,00	,01	-,01	,04	-,01
66	intexpy	Cost of funding earning assets	Performance and Condition Ratios	,00	,06	,45	-,08	-,01	-,15	-,02	,01	,04	,01	-,06	-,04	-,02	,02	-,01	-,19	,00
67	intincy	Yield on earning assets	Performance and Condition Ratios	,07	-,05	,38	,08	,00	-,10	-,01	-,01	,06	,01	,01	-,04	-,01	,00	-,02	,11	,01
68	lnatresr	Loss allowance to loans	Performance and Condition Ratios	,04	-,12	,00	-,07	,00	-,01	-,02	,01	-,05	,01	,05	,96	-,04	,00	,00	,00	,00
69	lnlsdepr	Net loans and leases to deposits	Performance and Condition Ratios	1,06	-,23	-,01	-,04	,00	-,02	,01	,00	,03	,00	-,02	,00	,00	,00	,00	-,01	,00
70	lnlsgr5	Average total loans	Performance and Condition Ratios	,70	,03	,01	-,08	,00	,54	-,02	-,01	,08	,00	,07	,01	,00	-,01	,00	,04	-,01
71	lnlsntv	Net loans and leases to total assets	Performance and Condition Ratios	,99	-,29	,05	-,02	,00	,02	-,01	,00	,34	,00	,20	-,03	,01	-,01	,00	-,10	,01
72	lnresncr	Loan loss allowance to noncurrent loans	Performance and Condition Ratios	-,02	,01	-,01	,01	,00	-,01	-,03	,00	,00	,00	-,03	,03	-,02	,02	,00	,01	,02
73	nclnlsr	Noncurrent loans to loans	Performance and Condition Ratios	,00	,01	,00	-,01	-,01	,00	,01	,00	-,01	-,01	,00	-,04	,99	,00	,00	,00	,00
74	nimy	Net interest margin	Performance and Condition Ratios	,04	-,03	-,15	-,03	,00	,02	,00	,00	,01	,00	,03	-,01	,00	,00	,00	1,00	,00
75	noijy	Net operating income to assets	Performance and Condition Ratios	,00	,01	-,02	-,01	,20	,00	,00	,00	,00	,77	-,01	-,02	,08	,00	,00	,00	,01
76	noniiay	Noninterest income to average assets	Performance and Condition Ratios	,02	,00	-,02	,00	,00	,00	,00	,02	,00	,00	,02	-,01	,00	,00	,98	,00	,00
77	nonixay	Noninterest expense to average assets	Performance and Condition Ratios	,08	-,03	,07	-,22	,02	-,09	,15	-,02	,01	,00	,08	-,02	,00	,02	-,01	-,01	,01
78	nperfv	Noncurrent assets plus other real estate owned	Performance and Condition Ratios	-,02	-,04	,03	,08	,02	-,03	-,04	,00	,04	,00	-,02	,40	,21	,00	,00	-,02	-,01
79	rbc1aaj	Core capital (leverage) ratio	Performance and Condition Ratios	,02	-,05	,01	,96	,01	-,07	-,04	,00	,04	,01	-,03	-,05	,00	,01	,00	,01	,00
80	rbc1rwaj	Tier 1 risk-based capital ratio	Performance and Condition Ratios	-,06	-,01	,00	1,02	,00	-,03	,02	,02	-,02	,01	,04	-,03	-,01	,00	,00	-,02	,01
81	rbcrwaj	Total risk-based capital ratio	Performance and Condition Ratios	,12	-,03	-,01	,94	,00	-,02	,02	,01	-,05	,01	,04	-,03	-,01	,00	,01	,00	,01
82	roa	Return on assets (ROA)	Performance and Condition Ratios	,00	-,01	-,02	-,03	,80	,01	,00	,00	,00	-,12	-,01	,01	,02	,00	,00	,00	,00
83	roaptx	Pretax return on assets	Performance and Condition Ratios	-,02	,04	,00	,04	,76	-,03	-,01	,00	,00	,00	-,01	,02	-,02	,00	,00	-,01	,03
84	roe	Return on Equity (ROE)	Performance and Condition Ratios	,02	-,02	-,02	-,05	,59	-,01	,01	,00	,01	-,11	-,01	-,01	,01	,00	,00	,00	,00
85	roeinjr	Retained earnings to average equity (ytd only)	Performance and Condition Ratios	,01	-,01	,01	-,02	,48	-,01	,01	,00	,02	-,03	,00	-,03	-,01	,01	,00	,00	-,01
86	scage	U.S. Government agency obligations	Securities	-,01	,07	,02	-,02	,00	-,02	,01	,61	,00	,00	-,01	-,01	,00	-,01	,00	,00	-,01
87	scrdebt	Total debt securities	Securities	,01	-,03	,00	,02	,00	,02	,00	,99	,02	,00	,02	,01	,00	,01	,00	,00	,00
88	scus	U.S. Government securities	Securities	-,03	,03	,00	-,04	,00	-,02	,00	,49	,00	,00	-,09	,00	,00	-,01	,00	,00	,01
89	ntrtmlg	Amount ($) - time deposits of $100,000	Time Deposits at the $100,000 Threshold	1,01	,00	-,03	-,06	-,01	-,04	-,01	,00	-,12	,01	,03	,02	,00	,01	,00	,02	,02
90	coredep	Retail deposits	Total Deposits	-,04	,60	-,06	-,03	,00	-,02	,05	,02	-,02	,00	,17	,05	,00	-,01	,00	,01	-,02
91	ddt	Demand deposits	Total Deposits	,00	-,01	-,01	,02	,00	-,02	,66	,00	,00	,00	,04	,03	,00	-,01	,00	,00	-,02
92	depidom	Interest-bearing deposits	Total Deposits	,03	,44	,00	,01	,00	,00	,09	,01	,00	,00	,01	,00	,00	,00	,00	,00	,78
93	depins	Estimated insured deposits	Total Deposits	-,04	,77	,08	-,03	,02	,01	-,04	-,01	,05	-,04	,19	,05	,00	-,03	,00	-,01	-,01
94	depnidom	Noninterest-bearing deposits	Total Deposits	-,03	,07	,02	,08	,00	-,01	,30	,00	,01	,00	,11	,01	,00	,00	,00	-,01	-,24
95	idtrni	Individuals, partnerships, and corporations	Total Deposits	-,10	,94	-,02	,02	,00	-,01	,04	-,01	,03	,00	,13	,00	,00	-,01	,00	,02	-,04
96	irakeogh	IRAs and Keogh plan accounts	Total Deposits	-,02	,03	-,02	,02	,00	,00	,00	,00	,00	,00	,00	,00	,00	-,02	,00	,00	,00
97	ntr	Nontransaction accounts	Total Deposits	-,06	,97	,01	-,02	,00	,01	-,16	-,01	,05	,00	,17	-,01	,00	,01	,00	-,01	,06
98	ntrsmmda	Money market deposit accounts (MMDAs)	Total Deposits	-,11	,33	-,01	,07	-,02	-,08	-,08	-,02	-,06	,01	-,12	,05	-,01	-,01	,00	,00	,12
99	ntrsoth	Other savings deposits (excluding MMDAs)	Total Deposits	-,03	,06	-,02	,00	,00	-,01	-,04	,00	,03	,01	-,01	,00	,00	,00	,00	,01	-,01
100	ntrtime	Total time deposits	Total Deposits	,77	,47	,00	-,09	,00	-,02	-,05	-,01	-,11	,00	,12	,00	,00	,01	,00	,01	,03
101	trn	Transaction accounts	Total Deposits	,03	-,02	,00	,01	,00	,00	1,01	,01	,00	,00	,09	-,03	,01	,00	,00	,00	,05
102	ts	Total time and savings deposits	Total Deposits	-,07	1,04	-,01	,01	,00	-,01	-,11	-,01	,00	,00	,22	-,01	,00	,00	,00	-,01	,09
103	edepdom	Interest expense: Domestic office deposits	Total Interest Expense	-,05	,03	1,00	,04	,00	,01	-,01	,00	-,01	,00	,05	,00	,00	,00	,00	-,03	,00
104	ilndom	Interest income: Domestic office loans	Total Interest Income	,19	-,03	,96	,03	,01	,01	-,01	,00	-,02	,00	,01	,00	,00	,00	,00	-,04	,00
105	isc	Interest income: Securities	Total Interest Income	-,07	,01	,03	-,09	,00	-,03	-,03	-,02	-,03	,00	-,25	,01	,00	-,01	,02	,00	,00
106	trnipcoc	Individuals, partnerships and corporations	Transaction Accounts	,04	-,11	,00	-,04	,00	,00	,97	,00	,02	,00	,01	-,03	,00	-,01	-,01	-,01	,06

FIGURE 3.20: Exploratory Factor Analysis: Scenario 2 (2 Part)

Scenario 3: Yearly changes in financial results

The data set of yearly changes in financial data meets the requirements of EFA. The measure of sampling adequacy is equal to 0.81, the Bartlett test is also significant (p=2.22e-16). The VSS test results are presented on the Figure 3.21. Total numbers of factors from 1 till 18 were tested during VSS.

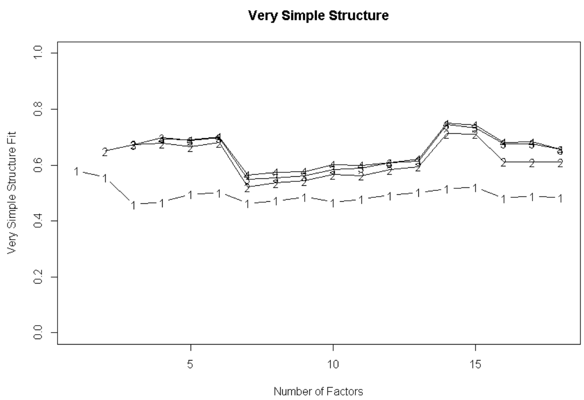


FIGURE 3.21: Very Simple Structure Fit: Scenario 3

The VSS achieves a maximum fit of 0.71 with 14 factors at complexity 2. Also, according to the test results, 14 factors provide the maximum average fit value of 0.75. Even though, the final number of factors is equal to 15, as the cumulative variance of 14 factors is not tolerable. The average fit of 15 factors is 0.74, which is a good result in comparison with remaining options. The cumulative variance explained by 15 factors is equal to 60.3%. The Kaiser criterion is satisfied, all eigenvalues are greater that 2, which means that each factor at least explains 2 exploratory variables. Cattell's scree test is also not useful as in the previous scenarios, as the total variance explained by the dimensions before the "last elbow" is too small to be acceptable. The loadings of factors are presented in the Figure 3.22 and Figure 3.23 The factors obtained from EFA with their interpretations are provided below:

- Factor 1: Represents Assets and Liabilities, Expense and Income related variables.
- Factor 2: Loans and Leases related variables.
- Factor 3: Income and Expense related variables.
- Factor 4: Deposits related financial data.
- Factor 5: Income (earnings).
- Factor 6: Equity (capital) results.
- Factor 7: Loans and Deposits.
- Factor 8: Total Assets and Total Liabilities.
- Factor 9: Interest income related variables.
- Factor 10: Non-transaction accounts.

- Factor 11: Non-interest Income related variables.
- Factor 12: Non-current loans.
- Factor 13: Operating income related data.
- Factor 14: Unused commitments.
- Factor 15: U.S. Government securities.

Name	Definition	Group	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15
1 eothnint	All other noninterest expense	Additional Noninterest Expense	,22	-,06	,34	,05	-,02	-,17	-,03	-,02	,02	-,02	-,02	-,01	-,02	,11	,17
2 asset	Total assets	Assets and Liabilities	-,01	-,03	,04	,01	,00	,03	-,01	1,00	,00	,00	,00	,00	,00	,00	,00
3 asset2	Average assets, quarterly	Assets and Liabilities	,37	-,03	-,01	-,01	,00	-,01	,02	,78	,00	-,01	,00	,01	,00	,00	-,01
4 asset5	Average total assets	Assets and Liabilities	,86	-,04	-,10	-,06	-,01	-,09	,03	,34	,00	,00	-,01	,01	,00	,00	-,01
5 AVASSETJ	Adjusted average assets for leverage capital purposes	Assets and Liabilities	,97	,00	-,12	,11	,00	-,08	,01	-,05	,01	-,03	,00	,03	,00	-,01	,00
6 bkprem	Bank premises and fixed assets	Assets and Liabilities	,05	,06	,00	,16	,00	,06	,01	,02	,00	-,02	,01	-,01	,00	,00	,00
7 chbal	Cash & Balances due from depository institutions	Assets and Liabilities	-,05	-,10	,03	,05	-,01	-,05	,07	,20	-,01	,00	-,01	,00	-,01	,00	,00
8 dep	Total deposits	Assets and Liabilities	,33	,00	-,05	,74	,01	-,01	-,02	,02	-,01	,01	,00	,00	,00	-,01	,00
9 depdom	Deposits held in domestic offices	Assets and Liabilities	,12	,00	-,01	,90	,01	,01	,02	,03	-,01	-,01	,00	,00	,00	-,01	-,01
10 depi	Interest-bearing deposits	Assets and Liabilities	,18	-,02	-,03	,59	,01	,02	-,09	,00	-,01	,01	,00	,00	,00	,00	-,01
11 eq	Bank equity capital	Assets and Liabilities	,27	,02	-,03	,04	,00	,47	-,02	,00	,00	-,01	,01	,00	,00	,01	,00
12 eqsur	Surplus	Assets and Liabilities	,01	,11	-,01	,19	-,01	,11	-,04	,07	-,02	-,01	-,01	,00	,00	,01	,01
13 eqtot	Total equity capital	Assets and Liabilities	,27	,02	-,03	,04	,00	,47	-,02	,00	,00	-,01	,01	,00	,00	,01	,00
14 equptot	Undivided profits	Assets and Liabilities	,01	-,01	,00	,01	,00	-,01	,01	,00	,00	-,01	,00	,00	,00	,00	,00
15 ernast	Earning assets	Assets and Liabilities	,74	,06	-,13	,29	,01	,01	-,01	,01	,01	-,02	,01	-,01	,00	,00	,00
16 idoa	All other assets	Assets and Liabilities	,40	,00	-,03	,05	-,02	-,05	,05	-,01	,00	,01	-,01	,04	-,01	-,03	,01
17 idoliab	All other liabilities	Assets and Liabilities	-,01	,02	,03	-,04	,00	-,01	,00	,49	,00	,00	,02	-,01	-,01	-,01	,03
18 liab	Total Liabilities	Assets and Liabilities	,00	-,02	,03	,02	,00	-,02	-,01	,99	-,01	,00	,00	,01	,00	,00	,00
19 liabeq	Total liabilities and capital	Assets and Liabilities	-,01	-,03	,04	,01	,00	,03	-,01	1,00	,00	,00	,00	,00	,00	,00	,00
20 lnatres	Loan loss allowance	Assets and Liabilities	-,15	-,01	1,03	,00	-,01	,02	,04	,02	,05	,00	,00	,04	,00	-,03	-,02
21 lnlsnet	Net loans and leases	Assets and Liabilities	,26	,78	-,04	,13	,00	-,02	,01	,00	,00	-,01	,00	-,01	,00	,00	,00
22 nclnls	Noncurrent loans and leases	Assets and Liabilities	,04	,00	,05	,00	,00	,01	,00	-,01	,00	,00	,00	,98	,00	,00	,00
23 numemp	Total employees (full-time equivalent)	Assets and Liabilities	,23	,07	,04	,10	,00	,05	-,05	,02	,00	-,04	,02	-,01	,00	,01	-,01
24 oaienc	Income earned, not collected on loans	Assets and Liabilities	,22	,02	,00	,16	,01	,00	,00	,00	,01	-,01	,01	,01	,01	,00	,00
25 RBCT1J	Tier one (core) capital	Assets and Liabilities	,38	,05	-,04	,08	,01	,78	,01	,01	,01	-,01	-,01	,00	,00	,01	,02
26 RBCT2	Tier 2 Risk-based capital	Assets and Liabilities	-,15	,01	1,02	,01	-,01	,08	,01	,03	,05	-,01	,01	,00	,00	-,04	-,04
27 RWAJT	Total risk weighted assets adjusted	Assets and Liabilities	,53	,40	-,09	,15	,01	-,02	-,03	,03	,00	-,03	,01	,02	,01	,00	,00
28 sc	Total securities	Assets and Liabilities	,05	-,05	-,02	,02	,00	,01	-,01	-,01	,01	,00	,00	-,01	,00	,00	,02
29 uc	Total unused commitments	Assets and Liabilities	-,01	,01	-,06	-,02	,00	,03	,01	,00	-,01	,01	,01	,00	,00	1,00	-,02
30 ucln	Unused loan commitments	Assets and Liabilities	-,01	,01	-,06	-,02	,00	,03	,01	,00	-,01	,01	,01	,00	,00	1,00	-,02
31 voliab	Volatile liabilities	Assets and Liabilities	,33	,03	-,09	-,07	,01	-,01	,25	,00	,00	,00	,00	,01	,01	,00	,00
32 LNRERSFM	Loans secured by 1-4 family first liens	1-4 Family Residential Net Loans and Leases	-,06	,34	,02	,03	,00	-,01	-,01	,04	-,01	,00	,00	,00	-,01	,00	,00
33 chbalni	Total noninterest-bearing balances	Cash and Balances Due	,00	-,03	,04	,03	,00	-,06	,14	,01	-,01	,00	,01	-,02	-,02	-,01	-,01
34 eqcprev	Amended balance at previous year-end	Changes in Bank Equity Capital	,45	-,03	,05	-,03	,00	,18	-,03	-,05	,00	,01	,00	-,02	-,01	,00	-,03
35 netinc	Net income	Changes in Bank Equity Capital	,00	,00	-,01	,00	,99	,00	,00	,00	,00	,00	,00	,00	,00	,00	,00
36 eintexp	Total interest expense	Income and Expense	,64	-,07	,23	-,08	,00	-,07	-,05	-,07	,08	,00	,00	-,01	,02	-,01	-,02
37 epremagg	Premises and equipment expense	Income and Expense	,15	,05	,31	,10	,00	-,01	,00	,00	,06	,02	,01	-,02	,00	-,01	-,02
38 esal	Salaries and employee benefits	Income and Expense	,22	,10	,34	,06	,02	,03	-,01	,00	,06	-,02	,02	-,02	-,01	,00	,02
39 ibefxtr	Income before extraordinary items	Income and Expense	,00	,00	-,01	,00	,99	,00	,00	,00	,00	,00	,00	,00	,00	,00	,00
40 IDEOTH	Additional noninterest expense	Income and Expense	,21	-,07	,31	,05	-,03	-,18	-,03	-,03	,02	-,01	-,11	-,01	-,02	,09	,13
41 idothnii	Additional Noninterest Income	Income and Expense	-,01	-,01	,01	,02	,00	-,02	,00	,00	,01	,00	,97	,00	,00	,01	,01
42 idpretx	Pre-tax net operating income	Income and Expense	-,01	,00	,04	,00	,09	-,01	,00	,01	,00	,00	-,03	,00	,81	,00	,00
43 intinc	Total interest income	Income and Expense	,16	,02	,14	-,04	,00	,02	,02	-,01	,94	,01	,01	,00	,00	-,01	-,01
44 nim	Net interest income	Income and Expense	,39	,00	,53	-,11	,01	,01	,03	-,03	-,45	,01	,02	-,01	,00	-,02	-,02
45 noij	Net operating income	Income and Expense	,00	,00	-,01	,00	-,04	,00	,00	,00	,00	,00	,01	,00	,99	,00	,00
46 nonii	Total noninterest income	Income and Expense	,00	-,01	,01	,00	,00	-,03	,00	,00	,01	,00	,99	,00	,00	,01	,01
47 nonix	Total noninterest expense	Income and Expense	,32	,02	,50	,04	,00	-,08	-,03	-,02	,05	-,02	-,02	-,02	-,02	,06	,04
48 lnls	Loans and leases, gross	Maturity & Repricing for Loans and Leases	,26	,78	-,03	,13	,00	-,03	,01	,00	,00	-,01	,00	,00	,00	,00	,00
49 idlnls	Loans and leases, gross	Net Loans and Leases	,26	,78	-,03	,13	,00	-,03	,01	,00	,00	-,01	,00	,00	,00	,00	,00
50 lnci	Commercial and industrial loans	Net Loans and Leases	,02	,02	,04	-,01	,00	,01	,04	,00	,00	,00	,01	,00	,00	,00	,00
51 lncon	Loans to individuals	Net Loans and Leases	,04	,01	,02	-,15	,00	,01	,90	-,01	,01	,02	-,01	,01	,00	,00	,01
52 lnconoth	Other loans to individuals	Net Loans and Leases	,02	,01	,01	-,07	,00	,02	,42	-,01	,01	,01	,00	,00	,00	,00	,00
53 lnlsgr	Total loans and leases	Net Loans and Leases	,26	,78	-,03	,13	,00	-,03	,01	,00	,00	-,01	,00	,00	,00	,00	,00

FIGURE 3.22: Exploratory Factor Analysis: Scenario 3 (1 Part)

Name	Definition	Group	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15
54 lnre	All real estate loans	Net Loans and Leases	,00	,81	,00	,08	-,01	-,01	-,01	,06	,00	,03	-,01	,00	-,01	,01	,00
55 lnredom	Real estate loans in domestic offices	Net Loans and Leases	,00	,81	,00	,08	-,01	-,01	-,01	,06	,00	,03	-,01	,00	-,01	,01	,00
56 lnrenres	Secured by nonfarm nonresidential properties	Net Loans and Leases	-,24	,58	,03	,02	-,01	,02	-,01	,07	-,01	,01	-,01	-,01	,00	-,01	,00
57 lnreres	1-4 family residential loans	Net Loans and Leases	-,07	,34	,02	,03	-,01	-,01	-,01	,04	-,01	,00	-,01	,00	-,01	,00	,00
58 ntripc	Individuals, partnerships, and corporations	Nontransaction Accounts	-,03	,01	,00	,00	,00	,01	,00	,00	,00	,99	,00	,00	,00	,00	,00
59 astempm	Assets per employee ($millions)	Performance and Condition Ratios	,44	-,10	-,12	,04	,00	-,11	,14	-,02	,00	,04	-,02	-,01	,00	-,03	,01
60 depdastr	Total domestic deposits to total assets	Performance and Condition Ratios	-,57	-,05	,12	,98	,00	,02	,05	-,17	-,02	,01	,00	,02	,00	,00	-,01
61 eeffr	Efficiency ratio	Performance and Condition Ratios	-,04	-,05	,00	,07	-,01	-,16	-,02	,01	-,02	-,01	-,04	-,01	-,01	,13	,00
62 eq5	Average equity	Performance and Condition Ratios	,57	,01	-,03	,00	-,01	,48	-,03	-,03	,01	,00	,00	-,03	,00	,02	-,01
63 eqv	Equity capital to assets	Performance and Condition Ratios	-,02	,03	,01	-,06	-,01	,47	-,01	-,01	,00	,00	,01	,02	,00	,01	,00
64 ernast5	Average earning assets	Performance and Condition Ratios	1,10	-,02	-,16	-,02	-,01	-,09	,04	-,09	,00	,00	,00	,01	,00	,00	-,01
65 idlncorr	Net loans and leases to core deposits	Performance and Condition Ratios	-,03	,11	,00	-,06	,00	,01	,01	,00	,00	,00	,02	,00	,00	,00	,00
66 intexpy	Cost of funding earning assets	Performance and Condition Ratios	,14	-,06	,53	-,07	,01	-,03	-,06	-,01	,12	-,01	,01	-,01	,03	-,03	-,02
67 intincy	Yield on earning assets	Performance and Condition Ratios	-,10	,01	,22	-,04	,01	,04	,01	,01	,96	,01	,01	,00	,00	-,01	-,01
68 lnatresr	Loss allowance to loans	Performance and Condition Ratios	-,25	-,10	1,04	,01	-,01	,01	,04	,02	,07	,01	,00	,04	,00	-,02	-,02
69 lnlsdepr	Net loans and leases to deposits	Performance and Condition Ratios	-,01	,91	,00	-,43	,01	,07	,01	-,05	,00	-,03	,00	-,01	,00	-,01	,00
70 lnlsgr5	Average total loans	Performance and Condition Ratios	,58	,57	-,04	-,05	,00	-,11	,01	-,06	-,01	,01	-,01	,02	,00	,00	,00
71 lnlsntv	Net loans and leases to total assets	Performance and Condition Ratios	-,29	1,04	,03	-,09	,00	,01	,01	-,12	,01	,01	,00	,01	,01	,00	,01
72 lnresncr	Loan loss allowance to noncurrent loans	Performance and Condition Ratios	-,28	-,01	,73	,03	,00	,08	,03	,03	,08	,00	,01	-,01	,00	-,03	-,03
73 nclnlsr	Noncurrent loans to loans	Performance and Condition Ratios	,02	-,01	,05	-,01	,00	,02	,00	,04	,00	,00	,00	,98	,00	,00	,00
74 nimy	Net interest margin	Performance and Condition Ratios	-,10	,02	,70	-,06	,01	,07	,03	,01	-,50	,01	,02	-,01	,00	-,02	-,02
75 noijy	Net operating income to assets	Performance and Condition Ratios	-,01	,00	-,01	,01	-,03	-,01	,00	,00	,00	,00	,01	,00	,99	,00	,00
76 noniiay	Noninterest income to average assets	Performance and Condition Ratios	,00	-,01	,01	,00	,00	-,03	,00	,00	,01	,00	,99	,00	,00	,01	,01
77 nonixay	Noninterest expense to average assets	Performance and Condition Ratios	-,31	,02	,65	,03	,02	-,05	-,01	,02	,07	-,01	-,02	-,02	-,02	,09	,06
78 nperfv	Noncurrent assets plus other real estate owned	Performance and Condition Ratios	,06	,01	,04	,01	,00	-,02	-,01	-,02	,00	,00	,00	,82	,00	,00	,00
79 rbc1aaj	Core capital (leverage) ratio	Performance and Condition Ratios	-,31	,11	,05	-,01	,00	1,00	,01	,03	,01	,01	-,02	,01	,00	,01	,02
80 rbc1rwaj	Tier 1 risk-based capital ratio	Performance and Condition Ratios	-,18	-,13	,05	,03	-,01	,95	,02	-,01	,01	,02	-,02	,00	-,01	,02	,02
81 rbcrwaj	Total risk-based capital ratio	Performance and Condition Ratios	-,17	-,15	,08	,01	-,01	,93	,01	-,01	,01	,02	-,02	,00	-,01	,01	,02
82 roa	Return on assets (ROA)	Performance and Condition Ratios	-,01	,00	-,01	,01	,99	,00	,00	,00	,00	,00	,00	,00	,00	,00	,00
83 roaptx	Pretax return on assets	Performance and Condition Ratios	,00	-,01	,00	,00	,87	,00	,00	,00	,00	,00	-,01	,00	-,04	,00	,00
84 roe	Return on Equity (ROE)	Performance and Condition Ratios	-,02	,00	,00	,00	,98	-,01	,00	,00	,00	,00	,02	,01	-,01	,00	,00
85 roeinjr	Retained earnings to average equity (ytd only)	Performance and Condition Ratios	-,02	,03	-,01	-,01	,17	,00	,00	,01	,00	,00	-,01	,01	,00	,00	,00
86 scage	U.S. Government agency obligations	Securities	-,03	,01	-,06	-,02	,00	,04	,00	,00	-,01	,00	,01	,00	,01	-,02	1,00
87 scrdebt	Total debt securities	Securities	,04	-,04	-,01	,01	,00	,02	-,01	-,01	,01	,00	,00	-,01	,00	,00	,02
88 scus	U.S. Government securities	Securities	-,03	,01	-,06	-,02	,00	,04	,00	,00	-,01	,00	,01	,00	,01	-,02	1,00
89 ntrtmlg	Amount ($) - time deposits of $100,000	Time Deposits at the $100,000 Threshold	,03	,07	-,01	,25	,00	-,06	,13	,02	,01	,13	,00	,01	,00	-,01	-,01
90 coredep	Retail deposits	Total Deposits	,10	,01	,01	,75	-,02	-,01	,06	,01	-,01	-,04	,00	-,02	,00	-,02	,01
91 ddt	Demand deposits	Total Deposits	,01	,01	,04	,00	,01	,01	,97	-,01	,01	,00	,00	,00	,00	,01	,00
92 depidom	Interest-bearing deposits	Total Deposits	,13	-,02	-,02	,60	,01	,01	-,08	,00	-,01	,06	,00	,00	,00	,00	-,01
93 depins	Estimated insured deposits	Total Deposits	-,15	-,03	,05	,80	,00	,02	,00	,01	-,02	-,05	,00	-,01	,00	-,01	-,01
94 depnidom	Noninterest-bearing deposits	Total Deposits	,04	,01	,04	-,01	-,01	,00	,94	-,01	,01	,01	,00	,00	,00	,01	,00
95 idtrni	Individuals, partnerships, and corporations	Total Deposits	,02	,02	,00	,88	,00	,06	,00	,03	-,02	,01	,00	,01	,00	,00	-,01
96 irakeogh	IRAs and Keogh plan accounts	Total Deposits	,23	,00	-,04	,09	,02	-,02	-,03	,00	,00	-,02	,00	-,01	,01	-,01	,00
97 ntr	Nontransaction accounts	Total Deposits	-,03	,00	,00	,02	,00	,01	,00	,00	,00	,98	,00	,00	,00	,00	,00
98 ntrsmmda	Money market deposit accounts (MMDAs)	Total Deposits	,09	-,01	-,02	,08	-,02	-,03	,00	-,01	,00	,00	-,03	-,01	-,01	-,01	,00
99 ntrsoth	Other savings deposits (excluding MMDAs)	Total Deposits	,08	-,01	,00	-,13	,00	-,01	-,01	-,01	,00	,68	,00	-,01	,00	,00	,00
100 ntrtime	Total time deposits	Total Deposits	,14	,03	-,01	,53	,00	-,06	-,07	,02	,00	,11	,01	,00	,00	-,01	,00
101 trn	Transaction accounts	Total Deposits	,05	-,02	,02	,14	,01	,01	,76	,01	,00	-,03	,01	-,01	,00	,00	,00
102 ts	Total time and savings deposits	Total Deposits	-,04	,00	,00	,02	,00	,02	,00	,00	,00	,82	,00	,00	,00	,00	,00
103 edepdom	Interest expense: Domestic office deposits	Total Interest Expense	,47	-,08	,27	,13	,01	-,01	-,09	-,07	,08	,03	,01	-,01	,02	-,01	-,02
104 ilndom	Interest income: Domestic office loans	Total Interest Income	,33	,33	,51	-,06	,00	-,03	,01	-,01	,10	,02	,01	-,01	,01	-,02	-,02
105 isc	Interest income: Securities	Total Interest Income	,00	-,02	-,08	-,02	,00	,00	,02	,00	,98	,00	,00	,00	,00	,00	,00
106 trnipcoc	Individuals, partnerships and corporations	Transaction Accounts	,01	,02	,02	,20	,00	,04	,72	,00	,00	-,02	,01	-,01	,00	,01	,00

FIGURE 3.23: Exploratory Factor Analysis: Scenario 3 (2 Part)

Chapter 4

Classification Experiments

This chapter presents the technical aspects and results of the experimental analysis of the available financial data using the implemented classification algorithms. The outputs of experiments within different data scenarios are described and compared. All analysis presented in this thesis were conducted in *RStudio IDE*.

4.1 Data Partition

In order to apply and assess the classification models, the data resulting from the dimensionality reduction and feature extraction processes were partitioned into two parts: training and testing data. Since the data sets have a sufficient size, the training part represents 70% of the total data and the testing part onsists of the remaining 30%. The data was partitioned randomly in order to mix the data from different time intervals and prevent separation of the specific time period. The data from the training part was used to build the classification model and cross-validation was used to tune the parameters of the supervised machine learning techniques. The testing data was then used to assess the prediction power of the classification models.

4.2 Cross Validation

Unbalanced data causes problems to many learning algorithms. These problems are characterized by the uneven proportion of cases that are available for each class of the problem. To avoid this problem, a hybrid approach of data sampling was used. The cross-validation results of the models based on the over-sampled minority class or the under-sampled majority class were compared to identify the optimal parameters of the classification model which can successfully predict both: bankrupt and non-bankrupt entities. The models tuning processes were conducted using the *train()* function from the *Caret* package.

4.3 The variables importance test

The measure of the importance of variables used in the analysis is a model-independent metric. Since there is no specific way to unambiguously assess the importance of each predictor, it is estimated individually using the "filter" approach. In our case, the receiver operating characteristic (ROC) curve analysis is calculated for each predictor. A series of cutoffs are applied to the predictor data to predict the class. Sensitivity and specificity are measured for each cutoff, on the basis of which the ROC curve is computed. The trapezoidal rule is used to approximate this computation. This area is used as the measure of variable importance. The importance of variables were assessed via *varImp()* function from the *Caret* package.

4.4 K-Nearest Neighbors

K-Nearest Neighbors (KNN) is one of the most commonly used and easy to understand classification algorithms in Machine Learning. The only parameter to be tuned is the value of *K*. The optimal *K* values were selected based on the accuracy of the model obtained from 5-fold cross validation using the over-sampled bankrupt class. The over-sampled data was used to prevent the impact of unbalanced data.

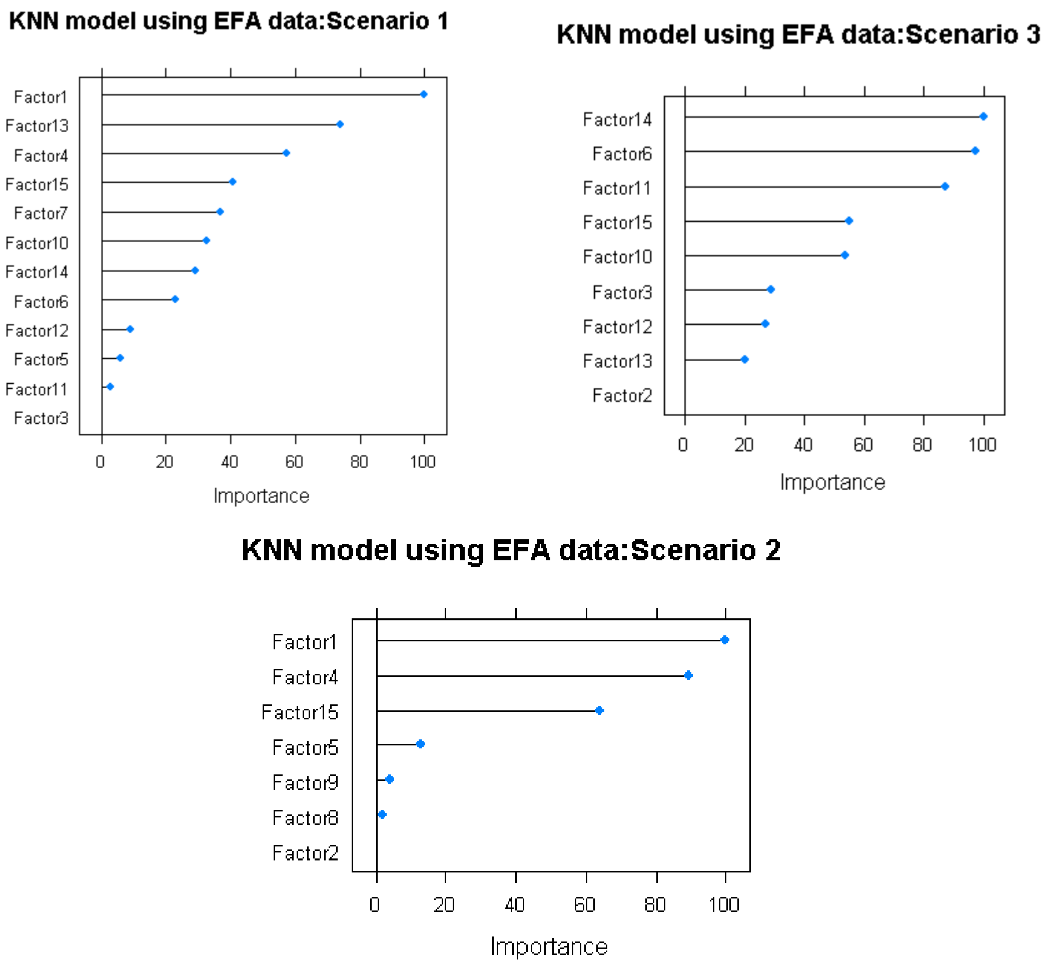


FIGURE 4.1: Importance of variables in KNN models obtained from EFA.

The KNN algorithm was applied on data obtained from EFA and PCA. In case of EFA, the removal of some factors increased the accuracy of the models in all observed scenarios. The best KNN model developed on quarterly data includes only 12 factors out of 17. The most important factor is the first factor which represents assets, liabilities, deposits, loans and leases related variables. For the model developed on the EFA data which represents the changes in financial results during half-year before expected bankruptcy, the most important variable is also factor number one with the same characteristics as in the quarterly data case. The model developed on the basis of EFA of yearly changes in financial results was mostly influenced from the factors which represent the changes in unused commitments and equity related variables. The summary of important variables in KNN models developed on EFA results is presented in Figure 4.1.

TABLE 4.1: KNN results

Period	Quarterly	Quarterly	Half-year	Half-year	Yearly	Yearly
Data	EFA	PCA	EFA	PCA	EFA	PCA
K	93	54	95	99	98	92
Accuracy	,911	,946	,927	,953	,898	,889
No Information Rate	,833	,833	,833	,833	,834	,834
P-Value [Acc > NIR]	6,44E-10	2,00E-16	1,38E-13	2,00E-16	1,61E-06	3,78E-05
Kappa	,689	,811	,739	,824	,637	,582
Sensitivity	,760	,868	,786	,812	,714	,616
Specificity	,942	,962	,955	,981	,934	,943
Balanced Accuracy	,851	,915	,871	,897	,824	,780

The accuracy of KNN models that were built from the PCA data tended to decrease if when some PCs were removed from analysis. So, in all three scenarios, the final models were developed based on the total available number of components.

The results for the KNN models are presented in Table 4.1. The accuracy of the KNN model built from PCA data outperformed the EFA-based one in the first and second scenarios. However, the results of KNN model using the EFA data provided better results than PCA for the yearly scenario. Overall, and according to the results of the analysis, the KNN model best identifies possible bankruptcy of the entities during prediction of the quarterly data.

4.5 Random Forest

The RF algorithm requires tuning only the *mtry* parameter which represents the number of variables tried at each split. This parameter was tuned by choosing the best value within the range from 2 till 10 by comparing the accuracy obtained from the 5-fold cross validated RF models. As the result the *mtry* value which leads to the highest accuracy was selected. The number of trees used in RF algorithms is 500.

The RF models which were developed based on the EFA data, as well as the models build on PCA data, best performed using the total available number of dimensions. The accuracy of the models substantially decreased during building models based on partial sampling of variables.

The importance of variables in RF models using EFA data in scenarios 1 and 2 are quite similar, in both cases the most important factors explain assets, liabilities, deposits, loans, leases and equity related variables. In scenario 3, the most important three factors represent the changes in the equity, unused commitments and non-interest Income.

According to the obtained results, it is obvious that the overall prediction accuracy of models developed on PCA results is higher than the models based on EFA. Only in the case of scenario 1, the sensitivity of the EFA data based RF model is higher than PCA based result. Based on the balanced accuracy value, We can say that the scenario 2 (half-year changes) best predicts financial distress.

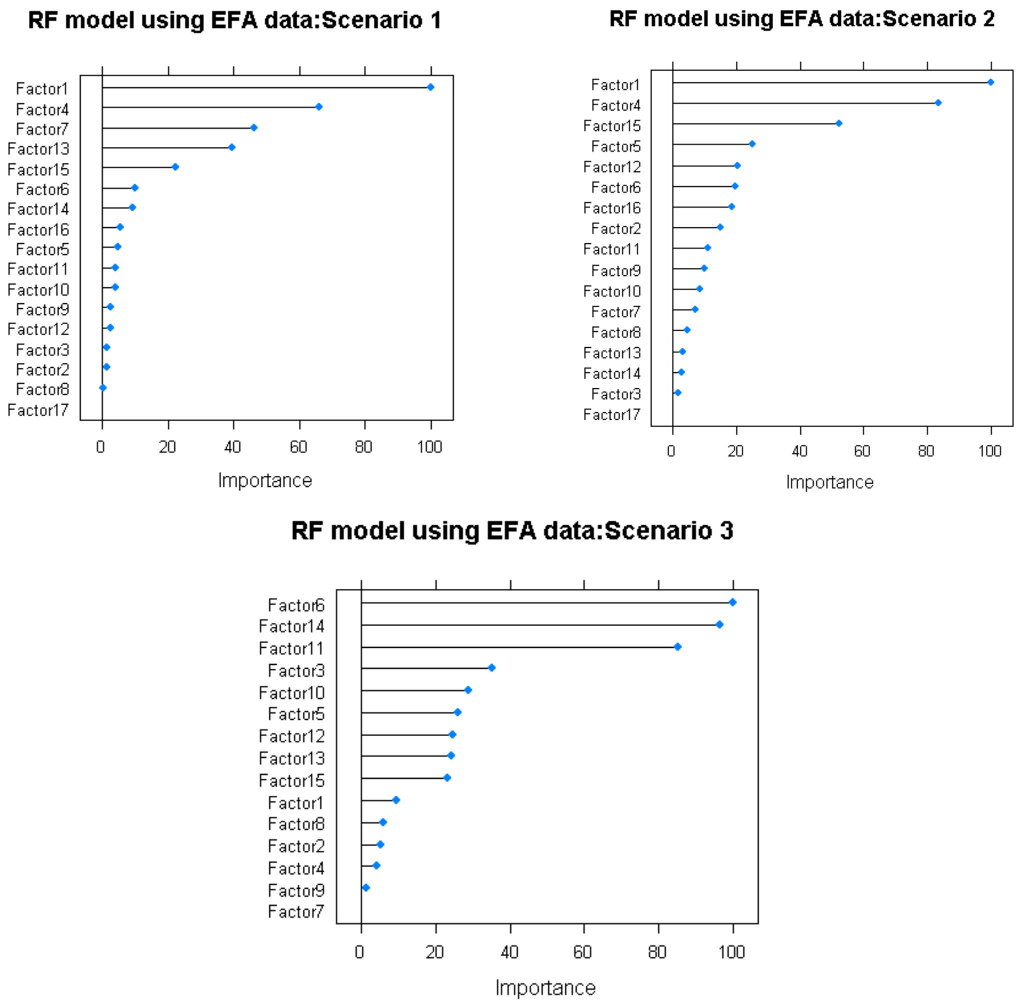


FIGURE 4.2: Importance of variables in RF models obtained from EFA.

TABLE 4.2: Random Forests results

Period	Quarterly	Quarterly	Half-year	Half-year	Yearly	Yearly
Data	EFA	PCA	EFA	PCA	EFA	PCA
mtry	4	9	3	7	3	10
Accuracy	.909	.957	.904	.913	.868	.871
No Information Rate	.833	.833	.833	.833	.834	.834
P-Value [Acc > NIR]	2.712e-09	2e-16	4.250e-08	6.376e-10	0.008555	0.004597
Kappa	.714	.844	.704	.732	.604	.616
Sensitivity	.909	.851	.914	.949	.857	.875
Specificity	.909	.978	.902	.906	.871	.871
Balanced Accuracy	.909	.915	.908	.927	.864	.873

4.6 Support Vector Machines

The SVM model was applied on the data obtained from the dimensionality reduction process. In SVM, the kernel function is applied to map the data in a higher dimensional space. In this study, the RBF kernel, two polynomial kernels, one with a degree of 2 and one with a degree 3, and the Gaussian kernel functions were examined. The analyses were conducted from the PCA and EFA resulting datasets.

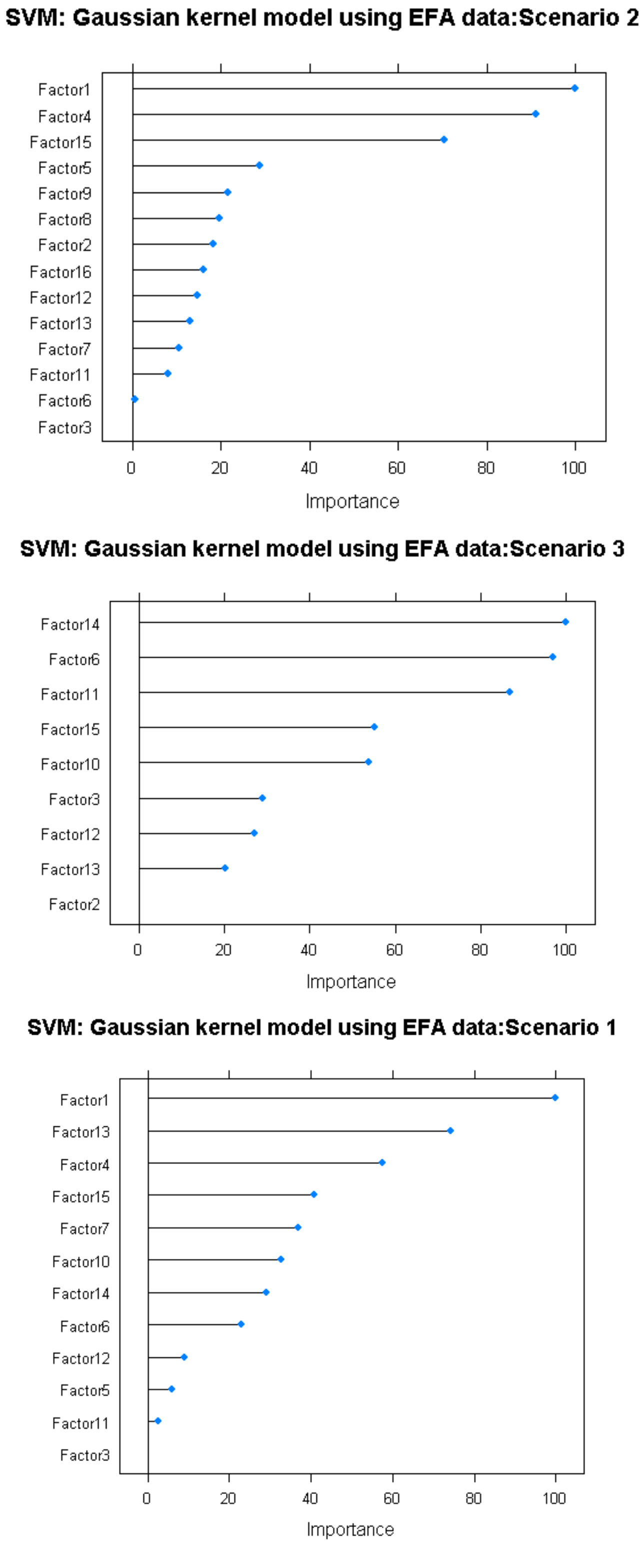


FIGURE 4.3: Importance of variables in Gaussian kernel SVM models obtained from EFA.

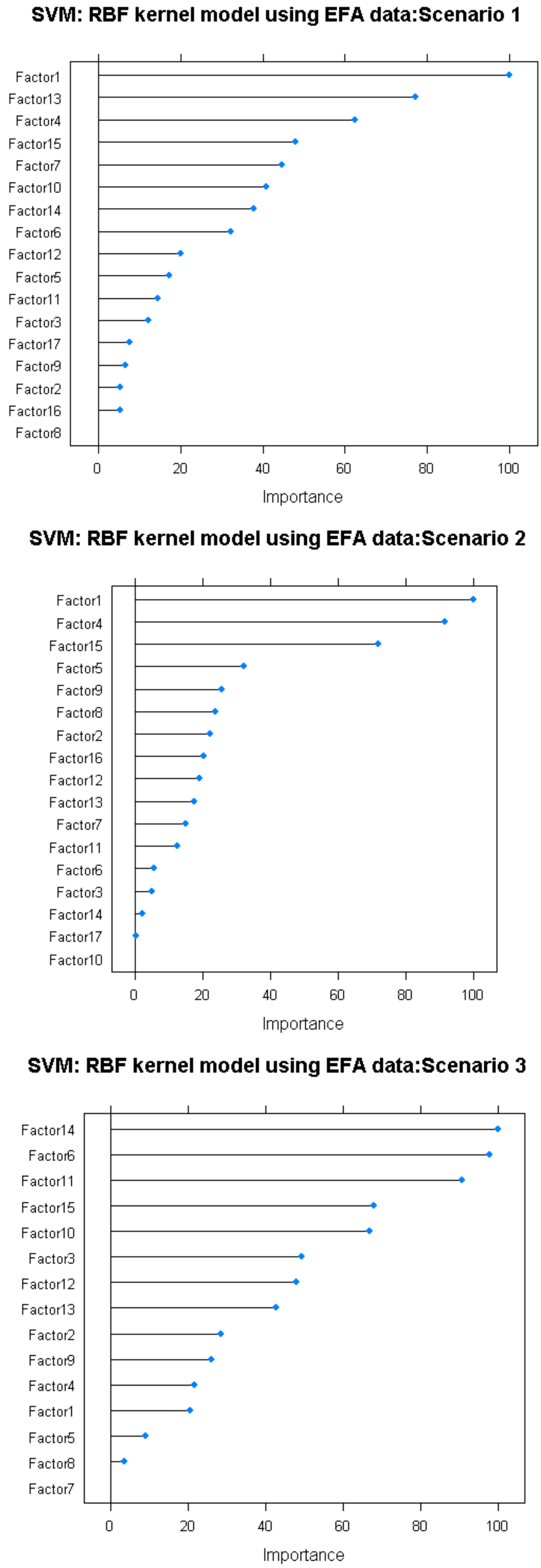


FIGURE 4.4: Importance of variables in RBF kernel SVM models obtained from EFA.

The parameters of the tested models were tuned by 5-fold cross validation. With the Gaussian kernel, one hyper-parameter needs to be tuned, i.e., σ. For RBF and the polynomial kernel, there is an additional parameter named "cost par

The parameters of the tested models were tuned by 5-fold cross validation. With the Gaussian kernel, one hyper-parameter needs to be tuned, i.e., σ. For RBF and the polynomial kernel, there is an additional parameter named "cost parameter", denoted by C, that requires tuning.

In all cases of the different scenarios, models were developed using the total available number of features. The RBF kernel SVM models were delevoped using both: the PCA and EFA data. The variables importance test obtained from the EFA data shows that the first two scenarios have same important factors which represent the changes in assets,liabilities, loans, leases, deposits and equity related variables. The third scenario results differ from the first two, in this case, the most important factors explain changes in unused commitments, assets, equity and deposits. These results are consistent with those previously presented in the KNN and RF models analysis.

The RBF kernel results outperformed all other kernels considered in this thesis in terms of sensitivity. Even if the total accuracy of the RBF models were least satisfactory related to others, the accuracy of financial distressed companies prediction was substantially higher than in other kernel methods. The hiighest sensitivity was obtained from RBF kernel SVM from the PCA data in the scenario 1, it equals to 96.7%. It is higher than the results obtained with RF and KNN models.

The Gaussian kernel performed best in the half-year data prediction using the EFA results. It provides 95.6% of total model accuracy and 92.6% of balanced accuracy. All the Gaussian models obtained from the EFA data provided better sensitivity results than PCA. The variables importance test obtained from the Gaussian models which were developed on the EFA data show that assets, liabilities, loans, leases, deposits, equity and U.S. Government securities-related factors are more important in the first scenario that the other features. The second scenario shows results of the importance test that are quite similar to the results of the scenario 1. In this case, assets, liabilities, loans, leases, deposits, equity and non-interest income related factors are estimated to be the most important. Scenario 3 shows completely different results in the importance test, as in this case unused commitments, equity and non-interest income-related variables are shown to be the most important.

The SVM with a 2-degree polynomial kernel provided better sensitivity and total accuracy in comparison with a 2 degree polynomial kernel. The Gaussian kernel performs best in the half-year data prediction using the EFA results. It provides 95.6% of total model accuracy and 92.6% of balanced accuracy. The importance test obtained from the SVM models with 2- and 3-degree polynomial kernels which were developed on the EFA data show exactly the same results, which are consistent with the results obtained with the previous models.

SVM models developed based on EFA provided better accuracy than PCA based models in observed scenarios 2 and 3. The models developed in scenario 1 show opposite results: in 3 out of 4 observed cases of SVM models, the accuracy of PCA based models were higher than EFA based models. Based on the obtained results, we can confidently say that the data corresponding to the changes of financial results during half-year before expected financial distress provide the best SVM classification results.

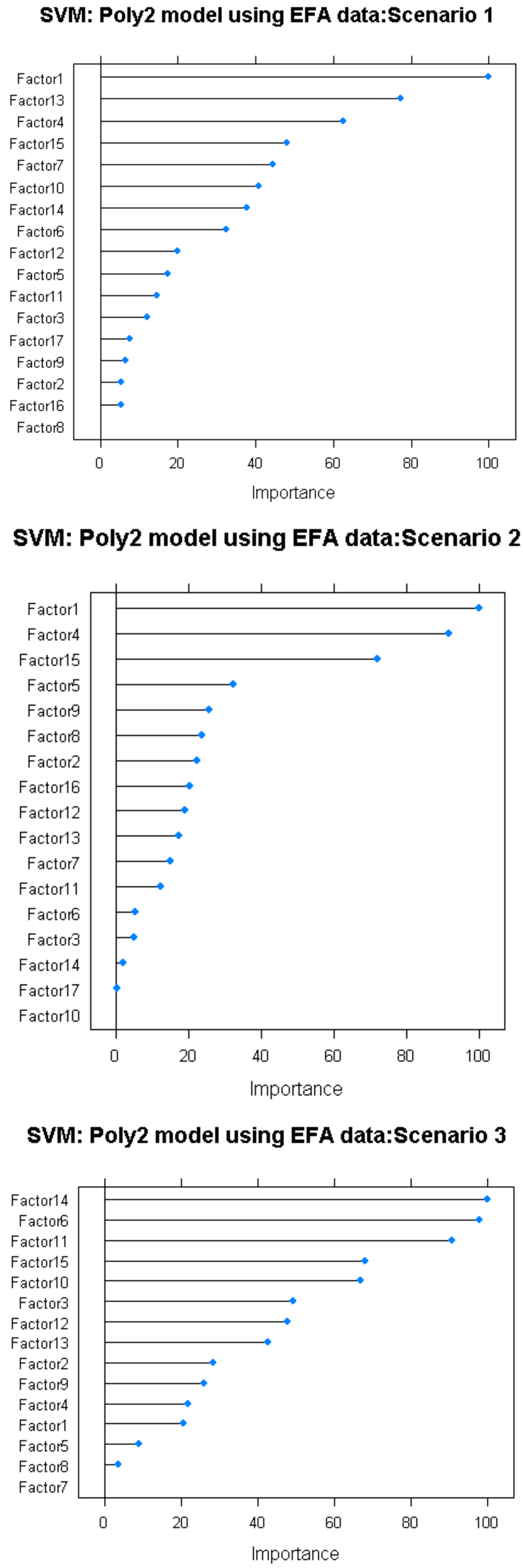


FIGURE 4.5: Importance of variables in 2-degree Polynomial kernel SVM models obtained from EFA.

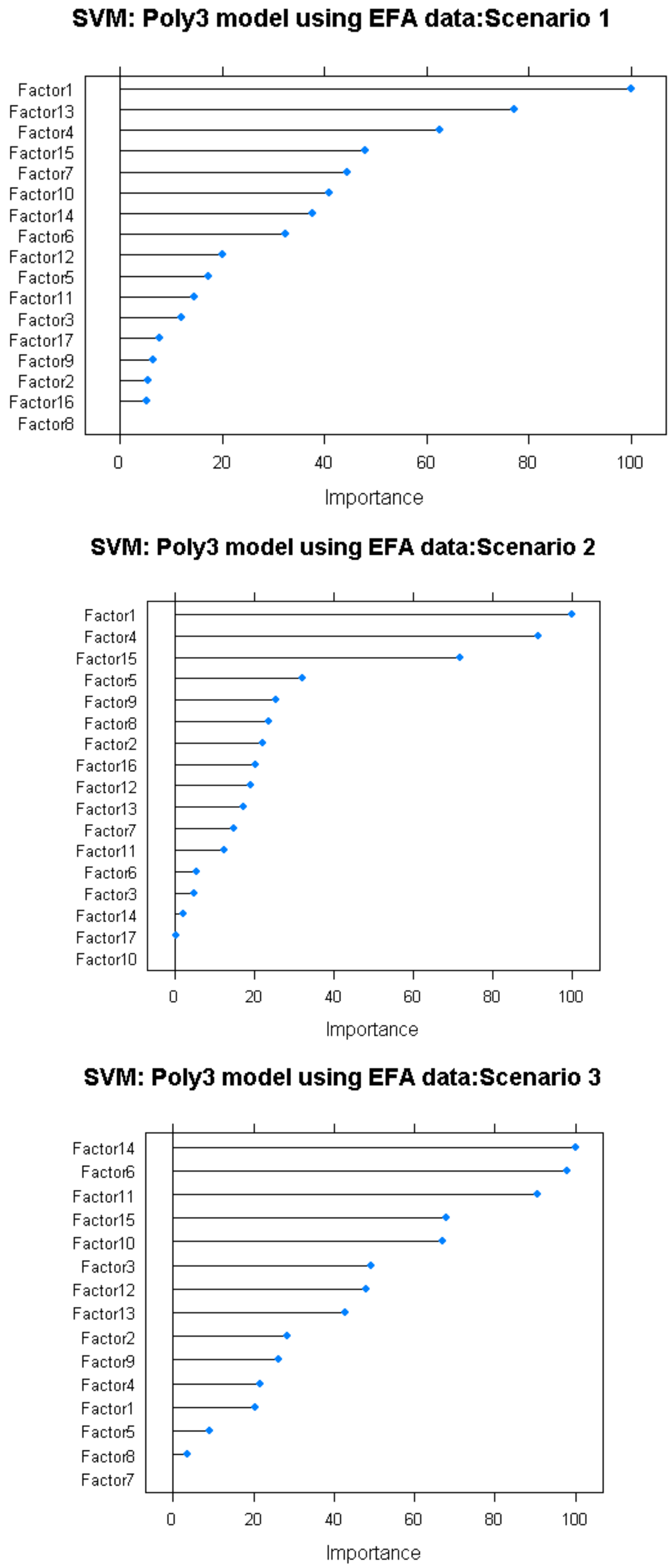


FIGURE 4.6: Importance of variables in 3-degree Polynomial kernel SVM models obtained from EFA.

TABLE 4.3: SVM results

Period	Quarterly	Quarterly	Half-year	Half-year	Yearly	Yearly
Data	EFA	PCA	EFA	PCA	EFA	PCA
	RBF	RBF	RBF	RBF	RBF	RBF
Accuracy	.895	.875	.907	.886	.859	.849
Sensitivity	.884	.967	.932	.949	.857	.839
Specificity	.897	.857	.902	.873	.860	.851
Balanced Accuracy	.891	.912	.917	.911	.858	.845
	Gaussian	Gaussian	Gaussian	Gaussian	Gaussian	Gaussian
Accuracy	.888	.943	.956	.953	.904	.882
Sensitivity	.818	.818	.880	.855	.777	.750
Specificity	.902	.968	.971	.973	.929	.908
Balanced Accuracy	.860	.893	.926	.914	.853	.829
	Poly-2	Poly-2	Poly-2	Poly-2	Poly-2	Poly-2
Accuracy	.920	.939	.949	.933	.898	.888
Sensitivity	.818	.926	.872	.914	.696	.696
Specificity	.940	.942	.964	.937	.938	.925
Balanced Accuracy	.879	.934	.918	.926	.817	.811
	Poly-3	Poly-3	Poly-3	Poly-3	Poly-3	Poly-3
Accuracy	.910	.927	.941	.934	.893	.899
Sensitivity	.777	.893	.846	.829	.670	.687
Specificity	.937	.934	.961	.955	.938	.941
Balanced Accuracy	.857	.913	.903	.892	.804	.814

4.7 Supervised SOM

The Supervised SOM method was used to develop a visually intuitive classification model for future financial prediction using the data obtained from the dimensionality reduction and feature extraction step. The analysis were performed using the XY-Fused networks algorithms which is one of the types of supervised SOM. My choice fell on XY-Fused networks algorithm, since it is a truly supervised approach, as discussed in the background information part of this study.

The supervised SOM has several parameters which should be selected for model development, this parameters include the size of X and Y dimensions, the topology of the map and weights given to individual layers. The hexagonal topology was used for map representation, all remaining parameters of supervised SOM were tuned by 5-fold cross-validation in order to avoid effect of unbalanced data. The models were trained by cross-validation and over- or under-sampling on each fold independently to get an honest estimate of model performance.

The visual outputs of the trained models are presented in Figure 4.8. The background colors of map units correspond to the trained models' classification results. The blue color represents the units which were classified as bankrupt, while green color represents non-bankrupt companies. The circle points on the maps represent the testing data points which were mapped into model units. The models performance presented on the plot may seem unsatisfactory, but, due to possible overlap of points in the map, it is impossible to judge the quality of the model based just on the plots. The plots were presented in order to show the principles of the supervised SOM. The real statistics of the developed supervised SOM models are presented in Table 4.4.

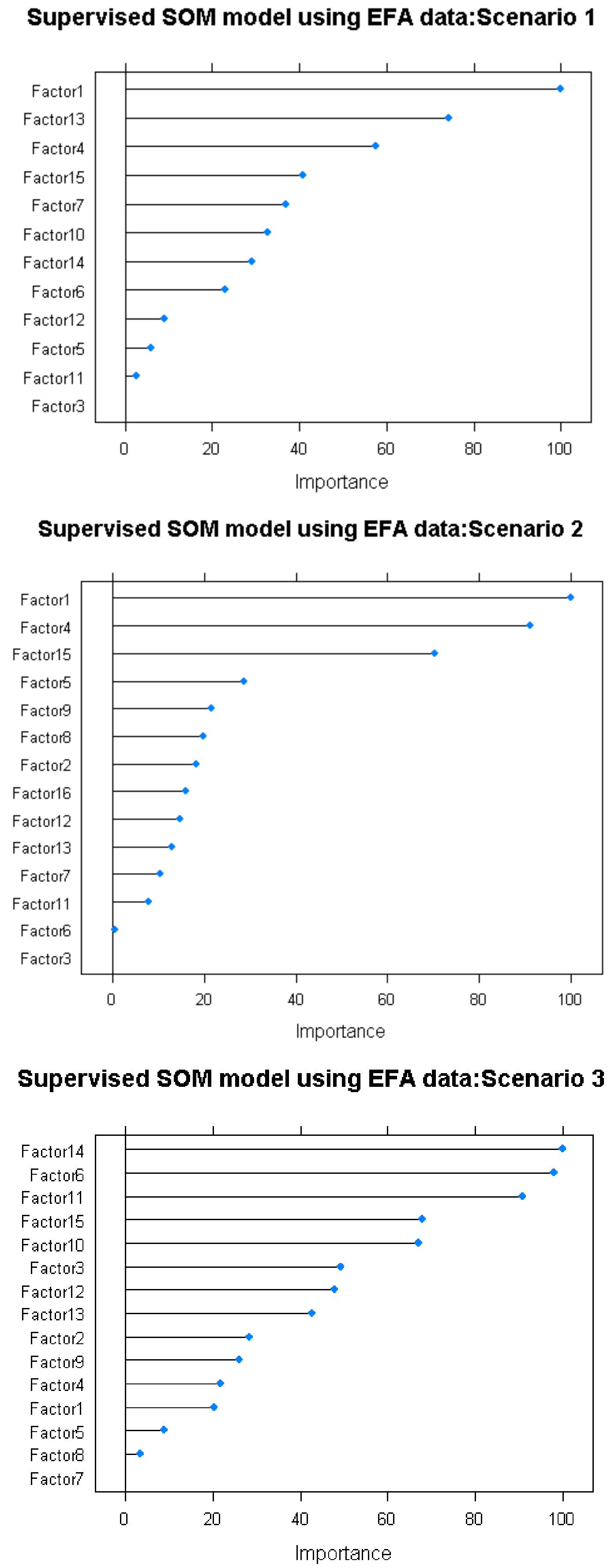


FIGURE 4.7: Importance of variables in supervised SOM models obtained from EFA.

The results of variables importance test is presented in Figure 4.7. Resetting of some variables from the analysis gave a sufficient improvement in accuracy results in EFA-based supervised SOM. The results obtained from the variables importance test match the results previously obtained in other classification models.

The supervised SOM analysis provides better results of sensitivities of models in all scenarios when the PCA are used to build the model. Using the changes of financial results for up to half-year before expected bankruptcy provides best prediction accuracy in supervised SOM case.

TABLE 4.4: Supervised SOM results

Period	Quarterly	Quarterly	Half-year	Half-year	Yearly	Yearly
Data	EFA	PCA	EFA	PCA	EFA	PCA
Size	5X5	4X4	3X3	5X5	3X3	4X4
Accuracy	.819	.934	.930	.934	.917	.800
No Information Rate	.833	.833	.833	.833	.834	.834
P-Value [Acc > NIR]	9.6e-10	4.816e-16	2.016e-14	9.546e-16	2.345e-10	1.696e-12
Kappa	.473	.768	.733	.769	.698	.443
Sensitivity	.752	.835	.718	.829	.741	.768
Specificity	.832	.953	.973	.955	.952	.807
Balanced Accuracy	.792	.894	.845	.892	.847	.787

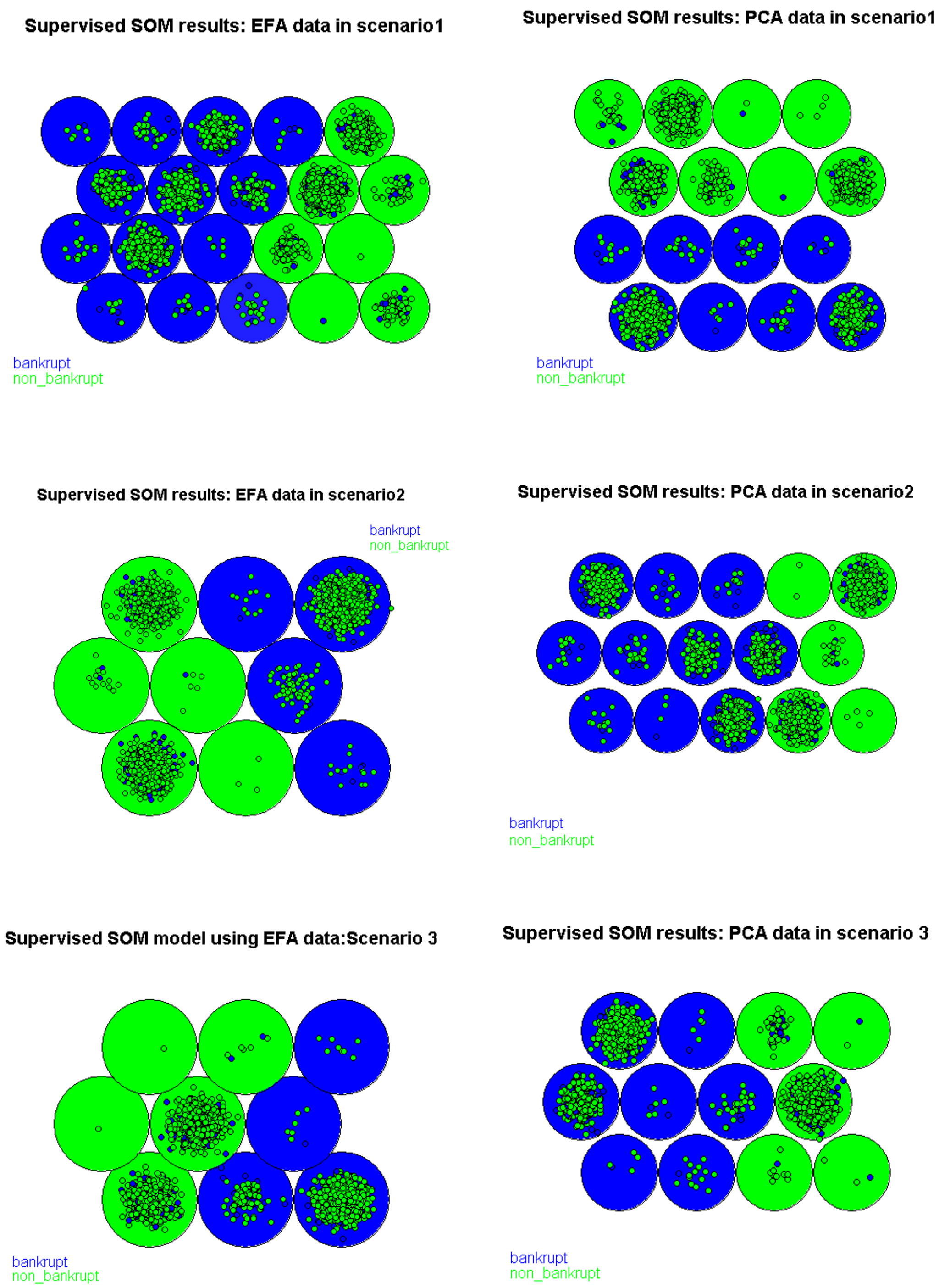


FIGURE 4.8: Supervised SOM models.

Chapter 5

Conclusion

The mission of this thesis was to build accurate financial distress prediction models from the historical data for three different periods previous to the expected bankruptcy, using different machine learning algorithms. The comparative review of different machine learning techniques like KNN, RF, SVM, and supervised SOM was conducted with the goal of identifying the best classification model among them.

All the classification models that were build based on the changes in historical financial data to predict the future financial distress in enterprises provided satisfactory results. This fact makes it possible to positively answer the original hypothesis, namely, that changes in historical financial data are relevant for the prediction of future financial problems in companies.

The comparison of observed machine learning techniques revealed that RBF kernel SVM models provides the best prediction accuracy of the companies under financial distress. The analysis were conducted under three different data sets. The first data set represents the changes in the financial data during a quarter prior to expected financial distress, the second data set represents the changes during a half-year prior to observed financial distress period and the third data set depicts the changes during the last year before expected bankruptcy day. The results of the classification methods obtained by using each of the data sets show, not unexpectedly, that the changes in historical financial data during a short term period, like quarter or half-year before bankruptcy, provide better results than predictions made using the changes during a year, but also that year-in-advance data still bear predictive power.

In this thesis, two dimensionality reduction and feature extraction techniques for data pre-processing were presented and tested in comparison to each other. One of them is PCA and the other is EFA. Using the data obtained from PCA most often yielded better classification results. However, as the number of data points increases, computational difficulties were encountered. The number of components retained after PCA is larger which greatly slows the computations compare to EFA. Compared with the EFA approach, the PCA results are harder to interpret in the input space. The EFA enabled extraction of the hidden information from the new features. The assets, liability, loans, leases, deposits and equity related factors had higher importance in the models build during the first two scenarios which represent quarterly and half-year changes in data. Assessing the importance of variables with the yearly data showed that unused commitments, assets and equity related factors have the highest importance.

There are several recommended directions for extending the present work. The first is to expand the observed dimensionality reduction techniques by the Kernel PCA and compare it with PCA and EFA. Secondly, to use the period-end financial results of existing data and try to build classification model based on them and compare the prediction results with the current work. Also, using the period end data implement Altman's Z-scores model and see if this approach is useful for the observed data's time interval, as some of the previous works revealed that this approach is not useful for the modern financial data. And, lastly, expand the observed data sets with the changes in financial data within longer time periods to check if the patterns identified in this work will be also relevant for them.

Bibliography

Adalessossi, Kokou (2015). “Prediction of Corporate Bankruptcy: Evidence from West African’s SMEs”. In: *Journal of Economics, Finance and Accounting – (JEFA)* 2.3.

Almendra, Vinicius and Denis Enachescu (2014). “Using Self-organizing Maps for Binary Classification with Highly Imbalanced Datasets”. In: *International Journal of Numerical Analysis and Modeling* 5.3, pp. 238–254.

Arora, Nidhi and JatinderKumar R. Saini (2014). “Bankruptcy Prediction of Financially Distressed Companies using Independent Component Analysis and Fuzzy Support Vector Machines”. In: *International Journal of Research in Computer and Communication Technology* 3.8. URL: https://www.researchgate.net/publication/281965431.

Back, B., G. Oosterom, and K. Sere (1994). “A comparative study of neural networks in bankruptcy prediction”. In: *In Proceedings of the 10th Conference on Artificial Intelligence Research*, pp. 140–148.

Bishop, Christopher M. (2006). *”Pattern Recognition and Machine Learning”*. URL: https://www.springer.com/kr/book/9780387310732.

Brown, Timothy A. (2006). *Confirmatory Factor Analysis for Applied Research”*. The Guilford Press, p. 486.

Buessler, Jean-Luc, Jean-Philippe Urban, and Julien Gresser (2002). “Additive Composition of Supervised Self-Organizing Maps”. In: *Neural Processing Letters* 15.9. URL: https://doi.org/10.1023/A:1013892727067.

Chen, Mu-Yen (2011). “Bankruptcy prediction in firms with statistical and intelligent techniques and a comparison of evolutionary computation approaches”. In: *Computers and Mathematics with Applications* 62, 4514–4524.

Crăciun, Mihaela et al. (2013). “Actuality of Bankruptcy Prediction Models used in Decision Support System”. In: *International Journal of Computers, Communications Control (IJCCC)* 8.3, pp. 375–383.

G., James et al. (2013). *”An Introduction to Statistical Learning with Applications in R”*. 1st ed. Springer-Verlag New York.

Hardle, Wolfgang and Leopold Simar (2003). *”Applied Multivariate Statistical Analysis”*. Springer-VerlagBerlinHeidelberg, p. 486. URL: www.i-xplore.de.

Hauser, Richard P. and David Booth (2011). “Predicting Bankruptcy with Robust Logistic Regression”. In: *Journal of Data Science* 9, pp. 565–584.

Hewson, Paul J. (2009). “Multivariate Statistics with R”. In: pp. 161–1.

Jones, Stewart (2016). “A Cash Flow Based Model of Corporate Bankruptcy in Australia”. In: *Journal of Applied Management Accounting Research - JAMAR* 14.1.

Kalyan, Nagaraj and Sridhar Amulyashree (2015). “A Predictive System for Detection of Bankruptcy Using Machine Learning Techniques”. In: *International Journal of Data Mining Knowledge Management Process* 5.1, pp. 29–40.

Kiviluoto, Kimmo (1998). “Predicting bankruptcies with the self-organizing map”. In: *Neurocomputing* 21, pp. 191–201.

Kline, Paul (1994). “An easy guide to factor analysis”. In: *Routledge*.

Klinke, Sigbert, Andrija Mihoci, and Wolfgang Hardle (2010). "Exploratory Factor Analysis in Mplus, R and SPSS". In: *ICOTS8*. URL: www.stat.auckland.ac.nz/~iase/publications.php.

Kohonen, Teuvo (2001). *Self-Organizing Maps*. 3rd ed. Springer-Verlag Berlin Heidelberg.

Kung, S. Y. (2014). *"Kernel Methods and Machine Learning"*. URL: http://www.cambridge.org/9781107024960.

Lantz, Brett (2013). *"Machine Learning with R"*. 2nd ed. Packt Publishing, p. 486.

Lukason, Oliver (2012). "Financial Performance Before Failure: Do Different Firms Go Bankrupt Differently?" In: *International Journal of Trade, Economics and Finance* 3.4.

Melssen, Willem, Ron Wehrens, and Lutgarde Buydens (2006). "Supervised Kohonen networks for classification problems". In: *Chemometrics and Intelligent Laboratory Systems* 83, 99–113.

Plucker, Jonathan (2003). "Exploratory and Confirmatory Factor Analysis in Gifted Education: Examples with Self-Concept Data". In: *Journal for the education of the gifted*. URL: https://www.researchgate.net/publication/285609869.

Ptak-Chmielewska, Aneta and Anna Matuszyk (2018). "The importance of financial and non-financial ratios in SMEs bankruptcy prediction". In: *Bank i Kredyt* 49.1, pp. 45–62.

Serrano-Cinca, Carlos (1996). "Self organizing neural networks for financial diagnosis". In: *Decision Support Systems* 17, pp. 227–238.

Thammano, Arit and Jiraporn Kiatwuthiamorn (2007). "Solving Classification Problems Using Supervised Self-Organizing Map". In: *2007 IEEE International Symposium on Signal Processing and Information Technology*, pp. 357–360.

Usuelli, Michele (2014). *"R Machine Learning Essentials"*. Packt Publishing.

Vapnik, Vladimir N. (1998). *"Statistical Learning Theory"*. 1st ed. Wiley-Interscience.

Vasighi, Mahdi and Mohsen Kompany-Zareh (2013). "Classification Ability of Self Organizing Maps in Comparison with Other Classification Methods". In: *MATCH Communications in Mathematical and in Computer Chemistry* 70, pp. 29–44. URL: https://www.researchgate.net/publication/236222198.

Václav Klepáč, David Hampel (2016). "Prediction of Bankruptcy with SVM Classifiers Among Retail Business Companies in EU". In: *Acta Univ. Agric. Silvic. Mendelianae Brun.* 64.2, pp. 627–634. URL: https://doi.org/10.11118/actaun201664020627.

Witten, Ian H., Frank Eibe, and Mark A. Hall (2011). *"Data mining : practical machine learning tools and techniques"*. 3rd ed. URL: https://www.elsevierdirect.com.

Wongravee, Kanet et al. (2010). "Supervised Self Organizing Maps for Classification and Determination of Potentially Discriminatory Variables: Illustrated by Application to Nuclear Magnetic Resonance Metabolomic Profiling". In: *Analytical Chemistry* 82.2, pp. 628–638.

Xiao, Yun-De et al. (2006). "Supervised Self-Organizing Maps in Drug Discovery. 2. Improvements in Descriptor Selection and Model Validation". In: *Journal of Chemical Information and Modeling* 46.1, pp. 137–144.

Appendix A

Descriptive statistics: quarterly data

	Name	Definition	GROUP	mean	median	mode	sd	min	max
1	eothnint	All other noninterest expense	Additional Noninterest Expense	,37	,46	1,00	,87	-,97	17,38
2	asset	Total assets	Assets and Liabilities	,00	,00	-,10	,06	-,66	,39
3	asset2	Average assets, quarterly	Assets and Liabilities	,00	,00	-,05	,05	-,34	,34
4	asset5	Average total assets	Assets and Liabilities	,00	,00	,00	,06	-,47	1,67
5	AVASSETJ	Adjusted average assets for leverage capital purposes	Assets and Liabilities	,00	,00	,00	,05	-,32	,35
6	bkprem	Bank premises and fixed assets	Assets and Liabilities	,01	-,01	,00	,15	-,81	3,94
7	chbal	Cash & Balances due from depository institutions	Assets and Liabilities	,23	,02	,00	1,52	-,90	54,86
8	dep	Total deposits	Assets and Liabilities	,01	,00	,00	,07	-,43	1,57
9	depdom	Deposits held in domestic offices	Assets and Liabilities	,01	,00	,00	,07	-,73	1,57
10	depi	Interest-bearing deposits	Assets and Liabilities	,01	,00	,00	,13	-1,00	5,03
11	eq	Bank equity capital	Assets and Liabilities	-,08	,01	,00	1,14	-25,88	36,77
12	eqsur	Surplus	Assets and Liabilities	,01	,00	,00	,10	-1,00	2,50
13	eqtot	Total equity capital	Assets and Liabilities	-,08	,01	,00	1,14	-25,88	36,76
14	equptot	Undivided profits	Assets and Liabilities	,01	,02	,00	2,14	-75,19	30,13
15	ernast	Earning assets	Assets and Liabilities	,00	,00	,00	,17	-,45	7,51
16	idoa	All other assets	Assets and Liabilities	,02	,00	,00	,33	-,83	9,77
17	idoliab	All other liabilities	Assets and Liabilities	,04	,00	,00	,53	-1,00	13,51
18	liab	Total Liabilities	Assets and Liabilities	,01	,00	-,10	,20	-,66	9,03
19	liabeq	Total liabilities and capital	Assets and Liabilities	,00	,00	-,10	,06	-,66	,39
20	lnatres	Loan loss allowance	Assets and Liabilities	,05	,01	,00	,25	-1,00	6,66
21	lnlsnet	Net loans and leases	Assets and Liabilities	-,01	,00	,00	,06	-,38	,38
22	nclnls	Noncurrent loans and leases	Assets and Liabilities	,57	,00	,00	5,58	-1,00	206,50
23	numemp	Total employees (full-time equivalent)	Assets and Liabilities	-,01	,00	,00	,09	-1,00	1,60
24	oaienc	Income earned, not collected on loans	Assets and Liabilities	,00	-,01	,00	,22	-,97	5,17
25	RBCT1J	Tier one (core) capital	Assets and Liabilities	-,09	,01	,00	,98	-25,88	20,64
26	RBCT2	Tier 2 Risk-based capital	Assets and Liabilities	-,03	,00	,00	,23	-4,25	1,12
27	RWAJT	Total risk weighted assets adjusted	Assets and Liabilities	,00	,00	,00	,07	-,80	,60
28	sc	Total securities	Assets and Liabilities	,02	,00	,00	,28	-1,00	6,46
29	uc	Total unused commitments	Assets and Liabilities	,09	-,02	,00	1,81	-1,00	70,89
30	ucln	Unused loan commitments	Assets and Liabilities	,08	-,02	,00	1,77	-1,00	70,89
31	voliab	Volatile liabilities	Assets and Liabilities	,04	,00	,00	1,23	-1,00	47,75
32	LNRERSFM	Loans secured by 1-4 family first liens	1-4 Family Residential Net Loans and Leases	,00	,00	,00	,17	-,94	6,57
33	chbalni	Total noninterest-bearing balances	Cash and Balances Due	,17	,00	,00	1,35	-,99	49,20
34	eqcprev	Amended balance at previous year-end	Changes in Bank Equity Capital	-,01	,00	,00	,16	-1,77	2,76
35	netinc	Net income	Changes in Bank Equity Capital	-,20	,34	,00	8,34	-266,18	35,00
36	eintexp	Total interest expense	Income and Expense	,25	,41	1,00	,65	-,91	1,90
37	epremagg	Premises and equipment expense	Income and Expense	,31	,46	1,00	,72	-,97	11,03
38	esal	Salaries and employee benefits	Income and Expense	,30	,46	1,00	,66	-,94	2,10
39	ibefxtr	Income before extraordinary items	Income and Expense	-,06	,34	,00	6,20	-208,25	35,00
40	IDEOTH	Additional noninterest expense	Income and Expense	,37	,46	1,00	,88	-,97	17,38
41	idothnii	Additional Noninterest Income	Income and Expense	,20	,37	1,00	3,69	-130,20	33,56
42	idpretx	Pre-tax net operating income	Income and Expense	,14	,35	,00	5,59	-208,25	76,57
43	intinc	Total interest income	Income and Expense	,28	,44	,00	,66	-,95	2,88
44	nim	Net interest income	Income and Expense	,30	,45	1,00	,69	-3,15	7,77
45	noij	Net operating income	Income and Expense	-,03	,35	,00	6,00	-208,25	33,00
46	nonii	Total noninterest income	Income and Expense	,28	,40	1,00	2,24	-60,50	33,56
47	nonix	Total noninterest expense	Income and Expense	,34	,47	,00	1,04	-,95	36,59
48	lnls	Loans and leases, gross	Maturity & Repricing for Loans and Leases	,00	,00	,00	,06	-,38	,35
49	idlnls	Loans and leases, gross	Net Loans and Leases	,00	,00	,00	,06	-,38	,35
50	lnci	Commercial and industrial loans	Net Loans and Leases	,00	-,01	,00	,20	-1,00	2,45
51	lncon	Loans to individuals	Net Loans and Leases	-,01	-,02	,00	,34	-1,00	14,00
52	lnconoth	Other loans to individuals	Net Loans and Leases	-,02	-,02	,00	,39	-1,00	14,00
53	lnlsgr	Total loans and leases	Net Loans and Leases	,00	,00	,00	,06	-,38	,35

	Name	Definition	Group	Mean	Median	Mode	Sd	Min	Max
54	lnre	All real estate loans	Net Loans and Leases	,00	,00	,00	,07	-,39	,72
55	lnredom	Real estate loans in domestic offices	Net Loans and Leases	,00	,00	,00	,07	-,39	,72
56	lnrenres	Secured by nonfarm nonresidential properties	Net Loans and Leases	,02	-,01	,00	,63	-1,00	28,68
57	lnreres	1-4 family residential loans	Net Loans and Leases	,00	,00	,00	,09	-,51	1,25
58	ntripc	Individuals, partnerships, and corporations	Nontransaction Accounts	,00	,00	,00	,07	-,51	,47
59	astempm	Assets per employee ($millions)	Performance and Condition Ratios	,01	,00	,00	,11	-,53	3,35
60	depdastr	Total domestic deposits to total assets	Performance and Condition Ratios	,01	,00	,00	,07	-,73	2,50
61	eeffr	Efficiency ratio	Performance and Condition Ratios	,07	,00	,00	1,39	-24,45	38,32
62	eq5	Average equity	Performance and Condition Ratios	-,04	,01	,00	,27	-3,26	7,74
63	eqv	Equity capital to assets	Performance and Condition Ratios	-,08	,00	,00	1,21	-27,88	38,56
64	ernast5	Average earning assets	Performance and Condition Ratios	,00	,00	,00	,05	-,53	,35
65	idlncorr	Net loans and leases to core deposits	Performance and Condition Ratios	-,01	-,01	,00	,10	-,63	2,64
66	intexpy	Cost of funding earning assets	Performance and Condition Ratios	-,05	-,03	,00	,10	-,57	2,52
67	intincy	Yield on earning assets	Performance and Condition Ratios	-,02	-,01	,00	,08	-,56	2,36
68	lnatresr	Loss allowance to loans	Performance and Condition Ratios	,05	,02	,00	,31	-1,00	10,07
69	lnlsdepr	Net loans and leases to deposits	Performance and Condition Ratios	-,01	-,01	,00	,07	-,63	,46
70	lnlsgr5	Average total loans	Performance and Condition Ratios	,00	,00	,00	,06	-,59	1,11
71	lnlsntv	Net loans and leases to total assets	Performance and Condition Ratios	,00	,00	,00	,07	-,34	1,83
72	lnresncr	Loan loss allowance to noncurrent loans	Performance and Condition Ratios	,37	,00	,00	3,99	-1,00	138,79
73	nclnlsr	Noncurrent loans to loans	Performance and Condition Ratios	,59	,00	,00	5,72	-1,00	202,68
74	nimy	Net interest margin	Performance and Condition Ratios	,00	,00	,00	,13	-2,11	3,58
75	noijy	Net operating income to assets	Performance and Condition Ratios	-,27	-,02	,00	3,95	-105,88	16,28
76	noniiay	Noninterest income to average assets	Performance and Condition Ratios	-,10	,00	,00	2,48	-66,41	22,07
77	nonixay	Noninterest expense to average assets	Performance and Condition Ratios	,03	,00	,00	,42	-,76	17,93
78	nperfv	Noncurrent assets plus other real estate owned	Performance and Condition Ratios	,31	,00	,00	4,56	-1,00	211,72
79	rbc1aaj	Core capital (leverage) ratio	Performance and Condition Ratios	-,10	,00	,00	1,04	-26,77	22,80
80	rbc1rwaj	Tier 1 risk-based capital ratio	Performance and Condition Ratios	-,09	,00	,00	1,08	-29,43	22,71
81	rbcrwaj	Total risk-based capital ratio	Performance and Condition Ratios	-,07	,00	,00	,80	-15,21	22,71
82	roa	Return on assets (ROA)	Performance and Condition Ratios	-,49	-,02	,00	7,72	-210,80	17,29
83	roaptx	Pretax return on assets	Performance and Condition Ratios	-,12	-,02	,00	4,58	-105,88	121,09
84	roe	Return on Equity (ROE)	Performance and Condition Ratios	-,38	-,01	,00	9,88	-218,29	107,97
85	roeinjr	Retained earnings to average equity (ytd only)	Performance and Condition Ratios	-,25	-,06	,00	10,71	-218,29	179,17
86	scage	U.S. Government agency obligations	Securities	,02	-,01	,00	,42	-1,00	12,30
87	scrdebt	Total debt securities	Securities	,02	,00	,00	,28	-1,00	6,46
88	scus	U.S. Government securities	Securities	,02	-,01	,00	,35	-1,00	6,46
89	ntrtmlg	Amount ($) - time deposits of $100,000	Time Deposits at the $100,000 Threshold	,03	,00	,00	,40	-1,00	11,42
90	coredep	Retail deposits	Total Deposits	,02	,01	,00	,10	-,73	1,73
91	ddt	Demand deposits	Total Deposits	,03	,00	,00	,28	-1,00	5,69
92	depidom	Interest-bearing deposits	Total Deposits	,01	,00	,00	,19	-1,00	6,96
93	depins	Estimated insured deposits	Total Deposits	,02	,01	,00	,10	-1,00	1,39
94	depnidom	Noninterest-bearing deposits	Total Deposits	,03	,01	,00	,41	-1,00	9,92
95	idtrni	Individuals, partnerships, and corporations	Total Deposits	,00	,00	,00	,09	-1,00	,42
96	irakeogh	IRAs and Keogh plan accounts	Total Deposits	,01	,00	,00	,20	-1,00	8,12
97	ntr	Nontransaction accounts	Total Deposits	,00	,00	,00	,06	-,50	,44
98	ntrsmmda	Money market deposit accounts (MMDAs)	Total Deposits	,04	,00	,00	,57	-,86	24,56
99	ntrsoth	Other savings deposits (excluding MMDAs)	Total Deposits	,02	,01	,00	,20	-1,00	5,79
100	ntrtime	Total time deposits	Total Deposits	,00	-,01	,00	,07	-,65	,61
101	trn	Transaction accounts	Total Deposits	,02	,00	,00	,20	-1,00	5,69
102	ts	Total time and savings deposits	Total Deposits	,01	,00	,00	,06	-,46	,44
103	edepdom	Interest expense: Domestic office deposits	Total Interest Expense	,25	,40	,00	,65	-,92	1,90
104	ilndom	Interest income: Domestic office loans	Total Interest Income	,29	,44	,00	,70	-,95	12,33
105	isc	Interest income: Securities	Total Interest Income	,27	,38	,00	,68	-1,00	7,14
106	trnipcoc	Individuals, partnerships and corporations	Transaction Accounts	,03	,00	,00	,44	-,88	17,64

Appendix B

Descriptive statistics: half-year data

	Name	Definition	Group	Mean	Median	Mode	Sd	Min	Max
1	eothnint	All other noninterest expense	Additional Noninterest Expense	,52	-,26	2,00	1,26	-1,16	7,77
2	asset	Total assets	Assets and Liabilities	,01	,01	-,22	,13	-,61	3,18
3	asset2	Average assets, quarterly	Assets and Liabilities	,01	,01	-,14	,10	-,54	2,64
4	asset5	Average total assets	Assets and Liabilities	,02	,01	-,09	,12	-,92	3,66
5	AVASSETJ	Adjusted average assets for leverage capital purposes	Assets and Liabilities	,02	,01	,00	,12	-,61	2,67
6	bkprem	Bank premises and fixed assets	Assets and Liabilities	,03	-,02	,00	,47	-1,00	13,98
7	chbal	Cash & Balances due from depository institutions	Assets and Liabilities	,40	,06	-,58	2,33	-1,00	73,11
8	dep	Total deposits	Assets and Liabilities	,02	,02	,00	,12	-1,00	2,64
9	depdom	Deposits held in domestic offices	Assets and Liabilities	,02	,02	,00	,12	-1,00	2,64
10	depi	Interest-bearing deposits	Assets and Liabilities	,02	,01	,00	,18	-1,00	4,36
11	eq	Bank equity capital	Assets and Liabilities	-,04	,02	,00	,51	-2,32	19,86
12	eqsur	Surplus	Assets and Liabilities	,15	,00	,00	3,18	-1,00	141,65
13	eqtot	Total equity capital	Assets and Liabilities	-,04	,02	,00	,51	-2,32	19,86
14	equptot	Undivided profits	Assets and Liabilities	-,32	,03	,00	11,11	-448,00	99,84
15	ernast	Earning assets	Assets and Liabilities	,01	,01	,00	,22	-,66	8,53
16	idoa	All other assets	Assets and Liabilities	,03	,00	,00	,26	-,95	4,63
17	idoliab	All other liabilities	Assets and Liabilities	,04	-,02	,00	,81	-21,80	23,04
18	liab	Total Liabilities	Assets and Liabilities	,02	,01	-,13	,11	-,92	2,63
19	liabeq	Total liabilities and capital	Assets and Liabilities	,01	,01	-,22	,13	-,61	3,18
20	lnatres	Loan loss allowance	Assets and Liabilities	,12	,03	,00	,56	-1,00	15,02
21	lnlsnet	Net loans and leases	Assets and Liabilities	,01	,00	,00	,59	-1,00	27,73
22	nclnls	Noncurrent loans and leases	Assets and Liabilities	1,62	,00	,00	21,60	-1,00	777,00
23	numemp	Total employees (full-time equivalent)	Assets and Liabilities	,00	,00	,00	,17	-,82	5,50
24	oaienc	Income earned, not collected on loans	Assets and Liabilities	-,01	-,04	,00	,55	-1,00	22,71
25	RBCT1J	Tier one (core) capital	Assets and Liabilities	-,05	,02	,00	,49	-3,70	19,88
26	RBCT2	Tier 2 Risk-based capital	Assets and Liabilities	,12	,00	,00	5,15	-1,00	249,01
27	RWAJT	Total risk weighted assets adjusted	Assets and Liabilities	,01	,00	,00	,14	-,72	3,11
28	sc	Total securities	Assets and Liabilities	,13	,00	,00	2,05	-1,37	63,45
29	uc	Total unused commitments	Assets and Liabilities	,08	-,02	,00	1,00	-1,00	34,22
30	ucln	Unused loan commitments	Assets and Liabilities	,08	-,02	,00	1,00	-1,00	34,22
31	voliab	Volatile liabilities	Assets and Liabilities	,06	,00	,00	2,24	-1,00	94,71
32	LNRERSFM	Loans secured by 1-4 family first liens	1-4 Family Residential Net Loans and Leases	,02	,00	,00	,41	-1,00	15,98
33	chbalni	Total noninterest-bearing balances	Cash and Balances Due	,36	,00	,00	3,93	-1,00	139,58
34	eqcprev	Amended balance at previous year-end	Changes in Bank Equity Capital	,01	,00	,00	2,20	-50,00	93,28
35	netinc	Net income	Changes in Bank Equity Capital	-,71	-,24	,00	31,37	-999,00	337,90
36	eintexp	Total interest expense	Income and Expense	,31	-,49	,00	1,08	-1,00	3,36
37	epremagg	Premises and equipment expense	Income and Expense	,43	-,39	2,00	1,17	-1,00	11,67
38	esal	Salaries and employee benefits	Income and Expense	,41	-,44	2,00	1,12	-,98	9,01
39	ibefxtr	Income before extraordinary items	Income and Expense	-,71	-,24	,00	31,37	-999,00	337,90
40	IDEOTH	Additional noninterest expense	Income and Expense	,54	-,26	2,00	1,48	-1,16	36,81
41	idothnii	Additional Noninterest Income	Income and Expense	,54	-,22	,00	9,00	-112,50	279,38
42	idpretx	Pre-tax net operating income	Income and Expense	-,89	-,27	,00	46,59	-2069,00	539,69
43	intinc	Total interest income	Income and Expense	,38	-,48	1,00	1,10	-1,00	3,52
44	nim	Net interest income	Income and Expense	,38	-,43	,00	1,57	-51,84	6,35
45	noij	Net operating income	Income and Expense	-1,29	-,25	-,33	50,02	-2069,00	210,19
46	nonii	Total noninterest income	Income and Expense	,91	-,34	,00	24,75	-109,14	1162,00
47	nonix	Total noninterest expense	Income and Expense	,45	-,41	,00	1,18	-1,00	14,76
48	lnls	Loans and leases, gross	Maturity & Repricing for Loans and Leases	,01	,00	,00	,55	-1,00	25,72
49	idlnls	Loans and leases, gross	Net Loans and Leases	,01	,00	,00	,55	-1,00	25,72
50	lnci	Commercial and industrial loans	Net Loans and Leases	,65	-,01	,00	29,08	-1,00	1405,92
51	lncon	Loans to individuals	Net Loans and Leases	-,02	-,03	,00	,38	-1,00	9,46
52	lnconoth	Other loans to individuals	Net Loans and Leases	-,04	-,04	,00	,38	-1,00	8,00
53	lnlsgr	Total loans and leases	Net Loans and Leases	,01	,00	,00	,55	-1,00	25,72

Name	Definition	Group	Mean	Median	Mode	Sd	Min	Max
54 lnre	All real estate loans	Net Loans and Leases	,01	,00	,00	,20	-1,00	5,15
55 lnredom	Real estate loans in domestic offices	Net Loans and Leases	,01	,00	,00	,20	-1,00	5,15
56 lnrenres	Secured by nonfarm nonresidential properties	Net Loans and Leases	,08	,00	,00	,91	-1,00	36,12
57 lnreres	1-4 family residential loans	Net Loans and Leases	,02	,00	,00	,35	-1,00	11,18
58 ntripc	Individuals, partnerships, and corporations	Nontransaction Accounts	,02	,01	,00	,13	-1,00	2,86
59 astempm	Assets per employee ($millions)	Performance and Condition Ratios	,03	,02	,00	,15	-,76	3,07
60 depdastr	Total domestic deposits to total assets	Performance and Condition Ratios	,01	,00	,00	,06	-1,00	,88
61 eeffr	Efficiency ratio	Performance and Condition Ratios	,15	,01	,00	3,79	-18,47	179,25
62 eq5	Average equity	Performance and Condition Ratios	-,02	,01	,00	,24	-2,16	4,57
63 eqv	Equity capital to assets	Performance and Condition Ratios	-,06	-,01	,00	,28	-3,06	3,99
64 ernast5	Average earning assets	Performance and Condition Ratios	,02	,01	,00	,13	-,98	3,65
65 idlncorr	Net loans and leases to core deposits	Performance and Condition Ratios	-21,56	-,03	,00	1042,36	-50422,52	28,87
66 intexpy	Cost of funding earning assets	Performance and Condition Ratios	-,11	-,09	,00	,15	-1,00	2,46
67 intincy	Yield on earning assets	Performance and Condition Ratios	-,05	-,03	,00	,09	-1,00	,39
68 lnatresr	Loss allowance to loans	Performance and Condition Ratios	,13	,03	,00	,62	-1,00	16,09
69 lnlsdepr	Net loans and leases to deposits	Performance and Condition Ratios	-,01	-,02	,00	,53	-,85	25,20
70 lnlsgr5	Average total loans	Performance and Condition Ratios	,01	,01	,00	,18	-,98	5,94
71 lnlsntv	Net loans and leases to total assets	Performance and Condition Ratios	-,01	-,01	,00	,16	-1,00	5,87
72 lnresncr	Loan loss allowance to noncurrent loans	Performance and Condition Ratios	,83	,00	,00	9,10	-1,00	261,52
73 nclnlsr	Noncurrent loans to loans	Performance and Condition Ratios	1,59	,01	,00	21,42	-1,00	778,52
74 nimy	Net interest margin	Performance and Condition Ratios	-,04	,00	,00	,50	-20,25	1,29
75 noijy	Net operating income to assets	Performance and Condition Ratios	-,53	-,03	,00	21,80	-668,37	347,88
76 noniiay	Noninterest income to average assets	Performance and Condition Ratios	,88	-,02	,00	48,74	-146,39	2346,36
77 nonixay	Noninterest expense to average assets	Performance and Condition Ratios	,03	,00	,00	,29	-,98	10,04
78 nperfv	Noncurrent assets plus other real estate owned	Performance and Condition Ratios	,79	,03	,00	7,90	-1,00	243,68
79 rbc1aaj	Core capital (leverage) ratio	Performance and Condition Ratios	-,07	-,01	,00	,28	-5,15	4,69
80 rbc1rwaj	Tier 1 risk-based capital ratio	Performance and Condition Ratios	-,06	,00	,00	,29	-5,40	4,08
81 rbcrwaj	Total risk-based capital ratio	Performance and Condition Ratios	-,06	,00	,00	,28	-5,28	5,90
82 roa	Return on assets (ROA)	Performance and Condition Ratios	-,05	-,03	,00	28,81	-504,11	1041,19
83 roaptx	Pretax return on assets	Performance and Condition Ratios	-,62	-,03	,00	13,39	-365,63	119,00
84 roe	Return on Equity (ROE)	Performance and Condition Ratios	,42	-,02	,00	48,63	-716,42	1547,61
85 roeinjr	Retained earnings to average equity (ytd only)	Performance and Condition Ratios	-,63	-,07	,00	39,34	-716,42	1231,41
86 scage	U.S. Government agency obligations	Securities	,23	-,01	,00	4,12	-1,00	145,60
87 scrdebt	Total debt securities	Securities	,13	,00	,00	1,98	-1,00	63,45
88 scus	U.S. Government securities	Securities	,31	-,01	,00	5,42	-1,00	169,46
89 ntrtmlg	Amount ($) - time deposits of $100,000	Time Deposits at the $100,000 Threshold	,08	,00	,00	1,52	-1,00	71,12
90 coredep	Retail deposits	Total Deposits	,04	,02	,00	,20	-1,00	7,09
91 ddt	Demand deposits	Total Deposits	,07	,01	,00	,69	-1,00	22,39
92 depidom	Interest-bearing deposits	Total Deposits	,02	,01	,00	,18	-1,00	4,36
93 depins	Estimated insured deposits	Total Deposits	,04	,02	,00	,15	-1,00	2,78
94 depnidom	Noninterest-bearing deposits	Total Deposits	,10	,02	,00	1,20	-1,00	41,15
95 idtrni	Individuals, partnerships, and corporations	Total Deposits	,02	,01	,00	,12	-1,00	2,53
96 irakeogh	IRAs and Keogh plan accounts	Total Deposits	,07	,01	,00	1,78	-1,00	85,87
97 ntr	Nontransaction accounts	Total Deposits	,02	,01	,00	,14	-1,00	3,18
98 ntrsmmda	Money market deposit accounts (MMDAs)	Total Deposits	,07	,00	,00	,65	-1,00	24,35
99 ntrsoth	Other savings deposits (excluding MMDAs)	Total Deposits	,11	,03	,00	1,76	-1,00	79,54
100 ntrtime	Total time deposits	Total Deposits	,01	,00	,00	,26	-1,00	9,45
101 trn	Transaction accounts	Total Deposits	,05	,02	,00	,40	-1,00	12,26
102 ts	Total time and savings deposits	Total Deposits	,02	,01	,00	,13	-1,00	2,72
103 edepdom	Interest expense: Domestic office deposits	Total Interest Expense	,31	-,49	,00	1,08	-1,00	5,49
104 ilndom	Interest income: Domestic office loans	Total Interest Income	,39	-,46	,00	1,15	-1,00	14,49
105 isc	Interest income: Securities	Total Interest Income	1,23	-,26	,00	41,23	-1,01	1994,00
106 trnipcoc	Individuals, partnerships and corporations	Transaction Accounts	,05	,02	,00	,48	-1,00	14,00

Appendix C

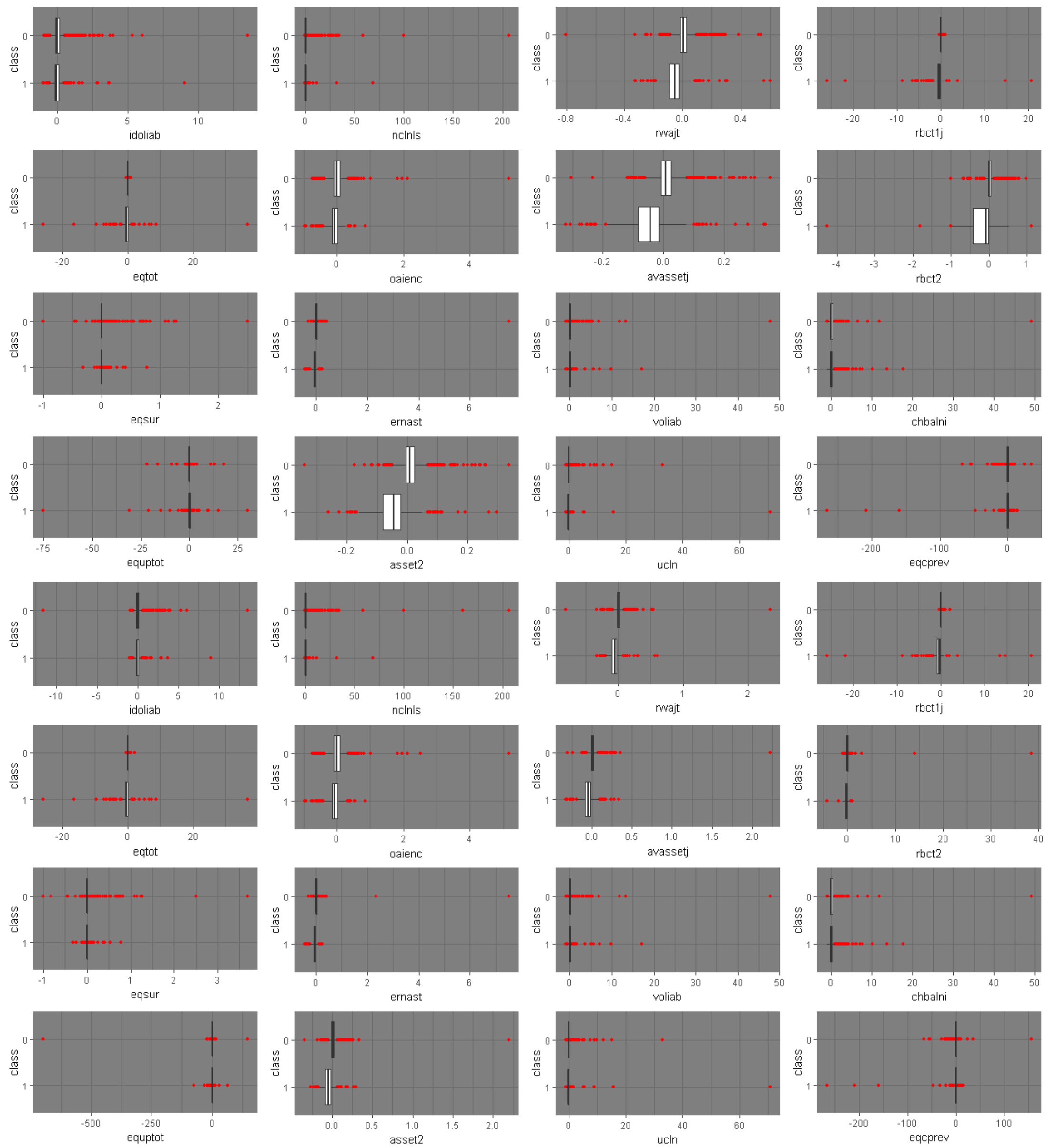
Descriptive statistics: yearly data

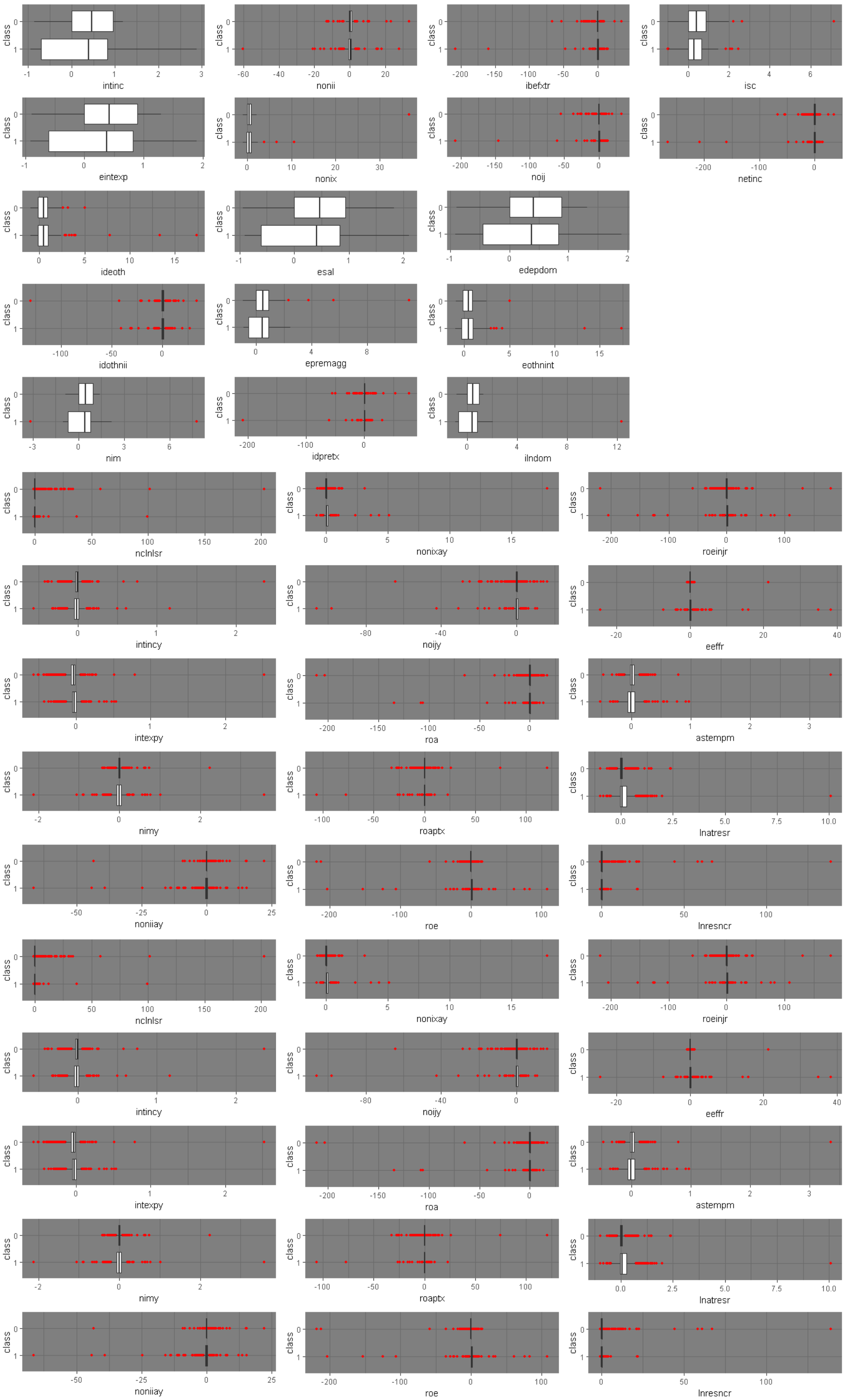
	Name	Definition	Group	Mean	Median	Mode	Sd	Min	Max
1	eothnint	All other noninterest expense	Additional Noninterest Expense	,17	,09	,00	,47	-,99	5,72
2	asset	Total assets	Assets and Liabilities	,06	,03	-,13	,37	-,96	15,55
3	asset2	Average assets, quarterly	Assets and Liabilities	,06	,04	-,09	,24	-,96	8,20
4	asset5	Average total assets	Assets and Liabilities	,06	,04	-,06	,18	-,96	3,27
5	AVASSETJ	Adjusted average assets for leverage capital purposes	Assets and Liabilities	,06	,04	,00	,17	-,96	1,82
6	bkprem	Bank premises and fixed assets	Assets and Liabilities	,09	-,03	,00	,76	-1,00	27,33
7	chbal	Cash & Balances due from depository institutions	Assets and Liabilities	,78	,12	,15	4,72	-,96	141,97
8	dep	Total deposits	Assets and Liabilities	,07	,04	,00	,22	-,96	3,21
9	depdom	Deposits held in domestic offices	Assets and Liabilities	,07	,04	,00	,24	-,96	4,72
10	depi	Interest-bearing deposits	Assets and Liabilities	,07	,03	,00	,36	-1,00	8,57
11	eq	Bank equity capital	Assets and Liabilities	,01	,04	,00	,47	-17,05	2,98
12	eqsur	Surplus	Assets and Liabilities	,22	,00	,00	2,47	-1,00	76,52
13	eqtot	Total equity capital	Assets and Liabilities	,01	,04	,00	,47	-17,05	2,98
14	equptot	Undivided profits	Assets and Liabilities	3,33	,05	,00	166,27	-731,00	7786,00
15	ernast	Earning assets	Assets and Liabilities	,05	,03	,00	,18	-,99	2,75
16	idoa	All other assets	Assets and Liabilities	,12	,04	,01	,50	-,96	11,22
17	idoliab	All other liabilities	Assets and Liabilities	,10	-,03	,00	,91	-,96	25,67
18	liab	Total Liabilities	Assets and Liabilities	,06	,04	-,12	,37	-,96	15,55
19	liabeq	Total liabilities and capital	Assets and Liabilities	,06	,03	-,13	,37	-,96	15,55
20	lnatres	Loan loss allowance	Assets and Liabilities	,37	,08	,00	2,94	-,99	109,00
21	lnlsnet	Net loans and leases	Assets and Liabilities	,05	,02	,00	,27	-,99	4,95
22	nclnls	Noncurrent loans and leases	Assets and Liabilities	4,80	,12	,00	51,46	-1,00	1995,00
23	numemp	Total employees (full-time equivalent)	Assets and Liabilities	,02	,00	,00	,33	-,96	12,47
24	oaienc	Income earned, not collected on loans	Assets and Liabilities	,02	-,04	,00	,63	-1,00	21,50
25	RBCT1J	Tier one (core) capital	Assets and Liabilities	,01	,03	,00	,24	-1,70	2,43
26	RBCT2	Tier 2 Risk-based capital	Assets and Liabilities	,23	,04	,00	2,92	-1,00	109,00
27	RWAJT	Total risk weighted assets adjusted	Assets and Liabilities	,05	,03	,00	,22	-,95	2,86
28	sc	Total securities	Assets and Liabilities	,30	,00	,00	4,31	-1,00	154,58
29	uc	Total unused commitments	Assets and Liabilities	,46	,00	,00	14,96	-1,00	703,60
30	ucln	Unused loan commitments	Assets and Liabilities	,45	,00	,00	14,95	-1,00	703,60
31	voliab	Volatile liabilities	Assets and Liabilities	,05	,00	,00	,87	-1,00	20,86
32	LNRERSFM	Loans secured by 1-4 family first liens	1-4 Family Residential Net Loans and Leases	,12	,02	,00	1,23	-1,00	49,21
33	chbalni	Total noninterest-bearing balances	Cash and Balances Due	,38	,00	,00	3,34	-1,00	121,22
34	eqcprev	Amended balance at previous year-end	Changes in Bank Equity Capital	,05	,04	,00	,33	-5,07	4,91
35	netinc	Net income	Changes in Bank Equity Capital	-,39	-,04	,00	13,41	-400,50	299,00
36	eintexp	Total interest expense	Income and Expense	-,10	-,16	,00	,48	-,99	15,28
37	epremagg	Premises and equipment expense	Income and Expense	,09	,02	,00	,50	-1,00	10,00
38	esal	Salaries and employee benefits	Income and Expense	,06	,03	,00	,41	-1,00	12,36
39	ibefxtr	Income before extraordinary items	Income and Expense	-,40	-,04	,00	13,44	-400,50	299,00
40	IDEOTH	Additional noninterest expense	Income and Expense	,19	,09	,00	,54	-1,08	7,00
41	idothnii	Additional Noninterest Income	Income and Expense	-,39	,01	,00	30,45	-1377,50	326,00
42	idpretx	Pre-tax net operating income	Income and Expense	,23	-,04	,00	14,01	-178,77	383,92
43	intinc	Total interest income	Income and Expense	,00	-,04	,00	,74	-,97	31,88
44	nim	Net interest income	Income and Expense	,05	,03	,00	,53	-12,60	9,98
45	noij	Net operating income	Income and Expense	,46	-,03	,00	28,43	-292,00	1220,90
46	nonii	Total noninterest income	Income and Expense	-,50	,01	,00	29,22	-1377,50	65,00
47	nonix	Total noninterest expense	Income and Expense	,09	,05	,00	,33	-1,00	5,76
48	lnls	Loans and leases, gross	Maturity & Repricing for Loans and Leases	,05	,03	,00	,27	-,99	4,96
49	idlnls	Loans and leases, gross	Net Loans and Leases	,05	,03	,00	,27	-,99	4,96
50	lnci	Commercial and industrial loans	Net Loans and Leases	,62	,00	,00	14,99	-1,00	602,36
51	lncon	Loans to individuals	Net Loans and Leases	,38	-,04	,00	14,78	-1,00	695,13
52	lnconoth	Other loans to individuals	Net Loans and Leases	,95	-,05	,00	32,29	-1,00	1362,00
53	lnlsgr	Total loans and leases	Net Loans and Leases	,05	,03	,00	,27	-,99	4,96

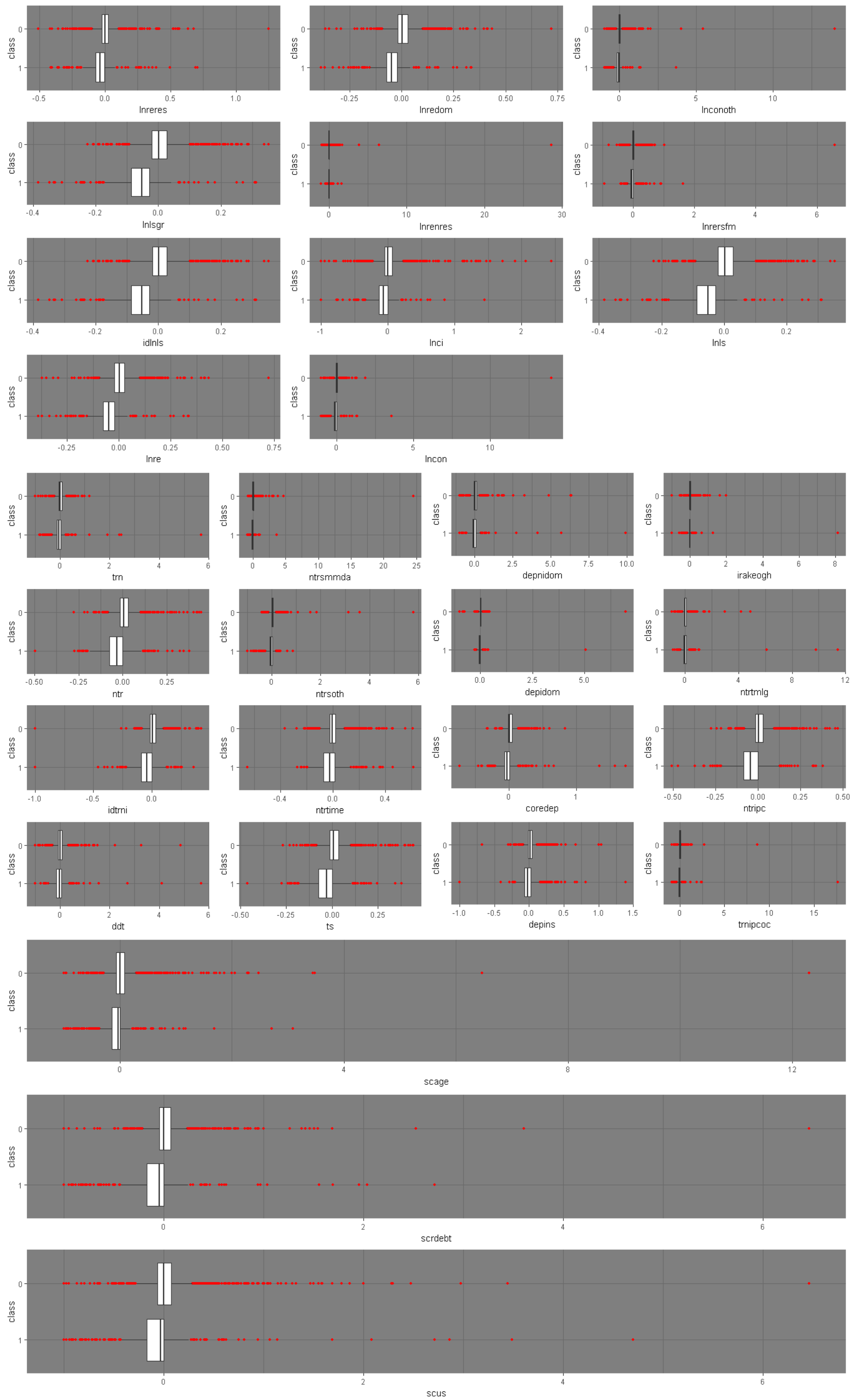
Name	Definition	Group	Mean	Median	Mode	Sd	Min	Max
54 lnre	All real estate loans	Net Loans and Leases	,07	,03	,00	,41	-,99	13,42
55 lnredom	Real estate loans in domestic offices	Net Loans and Leases	,07	,03	,00	,41	-,99	13,42
56 lnrenres	Secured by nonfarm nonresidential properties	Net Loans and Leases	,53	,02	,00	17,81	-1,00	843,64
57 lnreres	1-4 family residential loans	Net Loans and Leases	,10	,02	,00	1,14	-1,00	49,21
58 ntripc	Individuals, partnerships, and corporations	Nontransaction Accounts	,09	,03	,00	1,07	-1,00	48,14
59 astempm	Assets per employee ($millions)	Performance and Condition Ratios	,05	,04	,00	,18	-,90	2,69
60 depdastr	Total domestic deposits to total assets	Performance and Condition Ratios	,02	,00	,10	,14	-,94	3,96
61 eeffr	Efficiency ratio	Performance and Condition Ratios	,07	,02	,00	,70	-16,02	15,39
62 eq5	Average equity	Performance and Condition Ratios	,03	,04	,00	,25	-2,17	3,00
63 eqv	Equity capital to assets	Performance and Condition Ratios	-,04	-,01	,00	,47	-19,54	2,57
64 ernast5	Average earning assets	Performance and Condition Ratios	,06	,04	,00	,17	-,96	2,21
65 idlncorr	Net loans and leases to core deposits	Performance and Condition Ratios	,02	-,02	,00	1,51	-,72	65,03
66 intexpy	Cost of funding earning assets	Performance and Condition Ratios	-,15	-,19	,00	,30	-,98	7,88
67 intincy	Yield on earning assets	Performance and Condition Ratios	-,05	-,07	,00	,63	-,94	27,66
68 lnatresr	Loss allowance to loans	Performance and Condition Ratios	,30	,04	,00	2,51	-,99	107,66
69 lnlsdepr	Net loans and leases to deposits	Performance and Condition Ratios	-,01	-,02	,00	,15	-,78	2,09
70 lnlsgr5	Average total loans	Performance and Condition Ratios	,06	,03	,00	,26	-,66	5,46
71 lnlsntv	Net loans and leases to total assets	Performance and Condition Ratios	-,01	-,01	,00	,13	-,98	2,06
72 lnresncr	Loan loss allowance to noncurrent loans	Performance and Condition Ratios	,92	-,01	,00	12,22	-1,00	433,11
73 nclnlsr	Noncurrent loans to loans	Performance and Condition Ratios	4,31	,12	,00	40,08	-1,00	1303,91
74 nimy	Net interest margin	Performance and Condition Ratios	-,01	-,02	,00	,39	-11,11	9,52
75 noijy	Net operating income to assets	Performance and Condition Ratios	,41	-,08	,00	26,55	-281,13	1121,69
76 noniiay	Noninterest income to average assets	Performance and Condition Ratios	-,52	-,04	-1,00	27,66	-1304,41	68,40
77 nonixay	Noninterest expense to average assets	Performance and Condition Ratios	,03	,00	,00	,26	-,95	6,22
78 nperfv	Noncurrent assets plus other real estate owned	Performance and Condition Ratios	3,01	,14	,00	35,86	-1,00	1551,65
79 rbc1aaj	Core capital (leverage) ratio	Performance and Condition Ratios	-,04	-,01	,00	,19	-1,69	2,51
80 rbc1rwaj	Tier 1 risk-based capital ratio	Performance and Condition Ratios	-,03	-,01	,00	,21	-1,81	2,59
81 rbcrwaj	Total risk-based capital ratio	Performance and Condition Ratios	-,03	-,01	,00	,18	-1,62	2,36
82 roa	Return on assets (ROA)	Performance and Condition Ratios	-,39	-,09	,00	13,41	-399,95	295,99
83 roaptx	Pretax return on assets	Performance and Condition Ratios	-,29	-,10	,00	14,68	-399,95	295,99
84 roe	Return on Equity (ROE)	Performance and Condition Ratios	-,36	-,09	,00	16,40	-499,00	305,49
85 roeinjr	Retained earnings to average equity (ytd only)	Performance and Condition Ratios	-,30	-,17	,00	25,01	-691,61	510,85
86 scage	U.S. Government agency obligations	Securities	4,34	,00	,00	176,47	-1,00	8331,56
87 scrdebt	Total debt securities	Securities	,28	,00	,00	4,21	-1,00	154,58
88 scus	U.S. Government securities	Securities	4,35	-,01	,00	176,57	-1,00	8331,56
89 ntrtmlg	Amount ($) - time deposits of $100,000	Time Deposits at the $100,000 Threshold	,15	,04	,00	,98	-1,00	35,79
90 coredep	Retail deposits	Total Deposits	,09	,05	,00	,25	-,99	4,72
91 ddt	Demand deposits	Total Deposits	,11	,04	,00	,84	-1,00	34,38
92 depidom	Interest-bearing deposits	Total Deposits	,08	,03	,00	,39	-1,00	8,57
93 depins	Estimated insured deposits	Total Deposits	,11	,05	,00	,51	-,97	16,65
94 depnidom	Noninterest-bearing deposits	Total Deposits	,12	,05	,00	,86	-1,00	34,38
95 idtrni	Individuals, partnerships, and corporations	Total Deposits	,07	,03	,00	,27	-1,00	4,98
96 irakeogh	IRAs and Keogh plan accounts	Total Deposits	,11	,04	,00	,68	-1,00	24,01
97 ntr	Nontransaction accounts	Total Deposits	,10	,03	,00	1,07	-1,00	48,14
98 ntrsmmda	Money market deposit accounts (MMDAs)	Total Deposits	,19	,03	,00	1,21	-1,00	31,68
99 ntrsoth	Other savings deposits (excluding MMDAs)	Total Deposits	,17	,05	,00	1,61	-1,00	51,86
100 ntrtime	Total time deposits	Total Deposits	,06	,01	,00	,39	-1,00	11,79
101 trn	Transaction accounts	Total Deposits	,10	,05	,00	,59	-,99	17,41
102 ts	Total time and savings deposits	Total Deposits	,11	,03	,00	1,30	-1,00	48,14
103 edepdom	Interest expense: Domestic office deposits	Total Interest Expense	-,09	-,17	,00	,51	-,94	15,78
104 ilndom	Interest income: Domestic office loans	Total Interest Income	,01	-,03	,00	,36	-1,00	9,14
105 isc	Interest income: Securities	Total Interest Income	,17	-,06	,00	5,68	-1,79	264,00
106 trnipcoc	Individuals, partnerships and corporations	Transaction Accounts	,10	,04	,00	,62	-,99	17,41

Appendix D

Boxplots: quarterly data

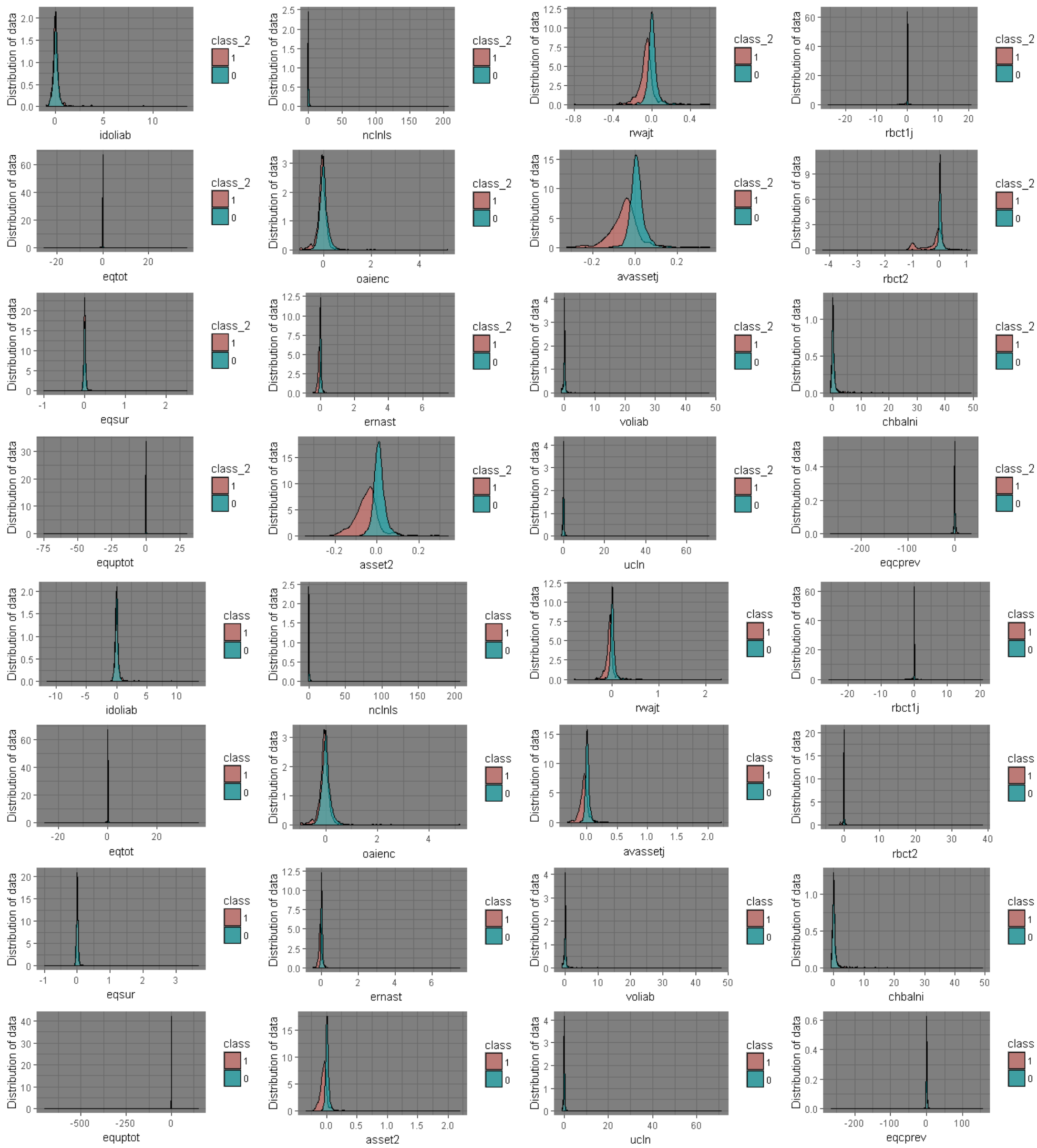


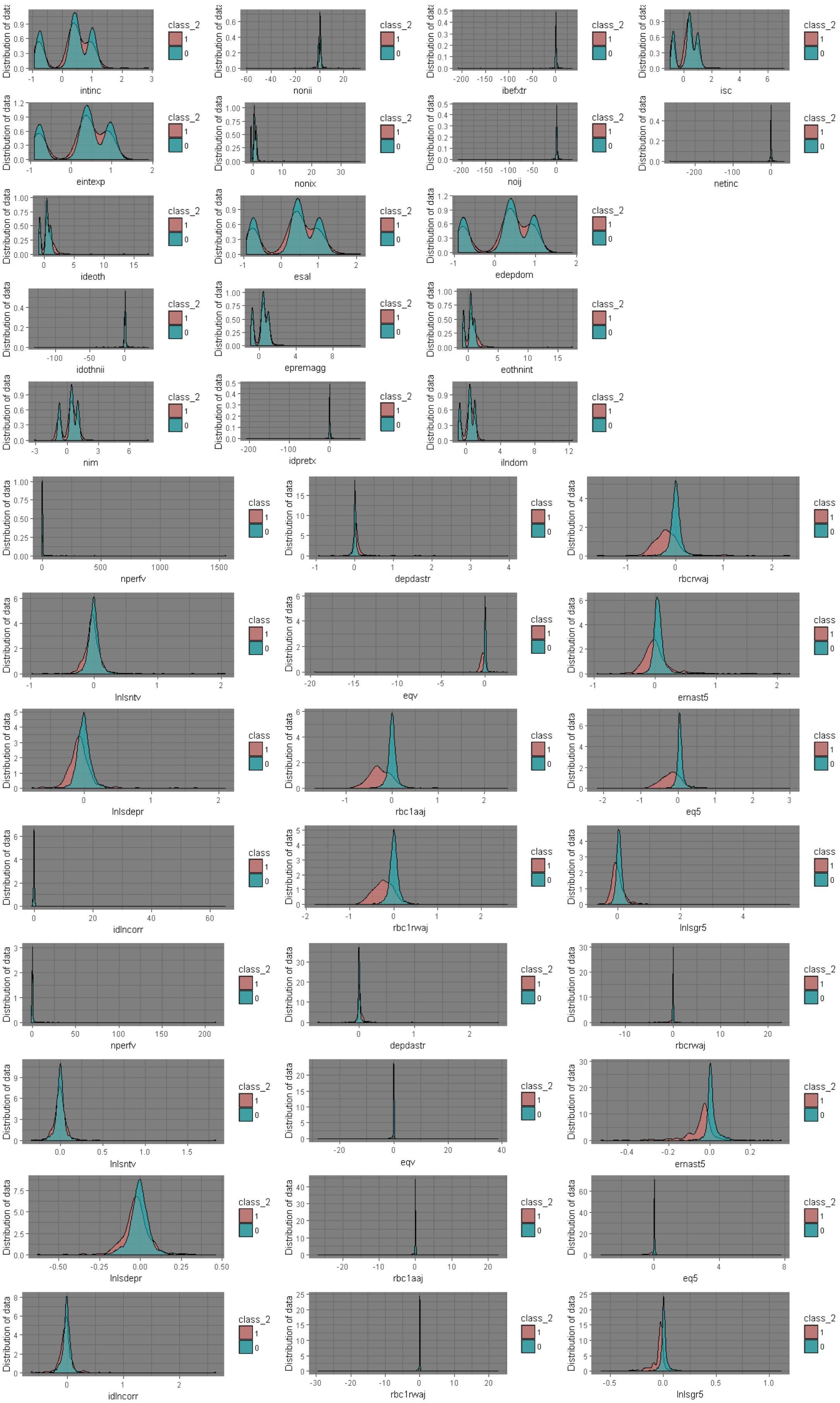

intinc
nonii
ibefxtr
isc
eintexp
nonix
noij
netinc
ideoth
esal
edepdom
idothnii
epremagg
eothnint
nim
idpretx
ilndom
nclnlsr
nonixay
roeinjr
intincy
noijy
eeffr
intexpy
roa
astempm
nimy
roaptx
lnatresr
noniiay
roe
lnresncr
nclnlsr
nonixay
roeinjr
intincy
noijy
eeffr
intexpy
roa
astempm
nimy
roaptx
lnatresr
noniiay
roe
lnresncr

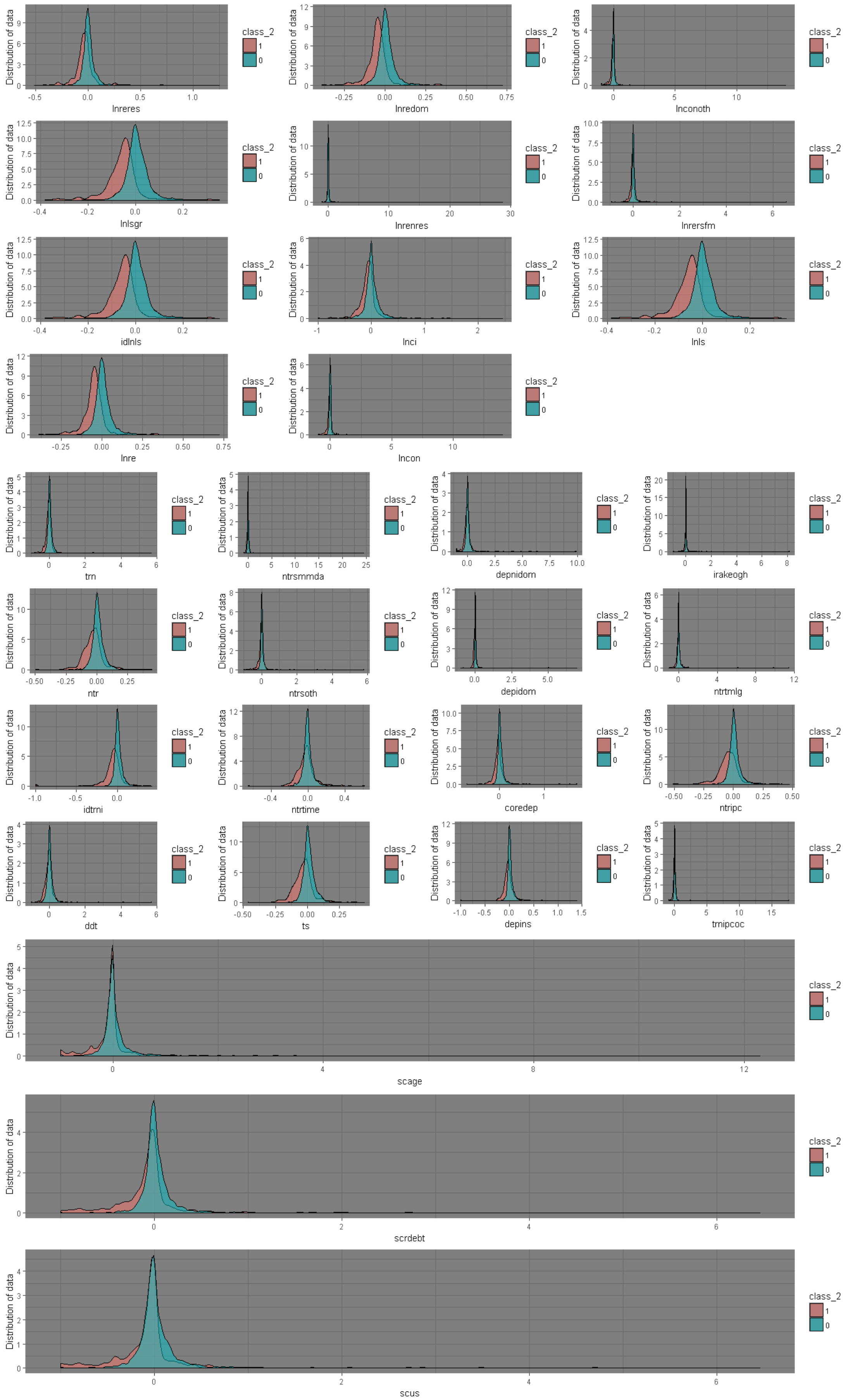

lnreres
lnredom
lnconoth
lnlsgr
lnrenres
lnrersfm
idlnls
lnci
lnls
lnre
lncon
trn
ntrsmmda
depnidom
irakeogh
ntr
ntrsoth
depidom
ntrtmlg
idtrni
ntrtime
coredep
ntripc
ddt
ts
depins
trnipcoc
scage
scrdebt
scus
class

Appendix E

Density plots: quarterly data

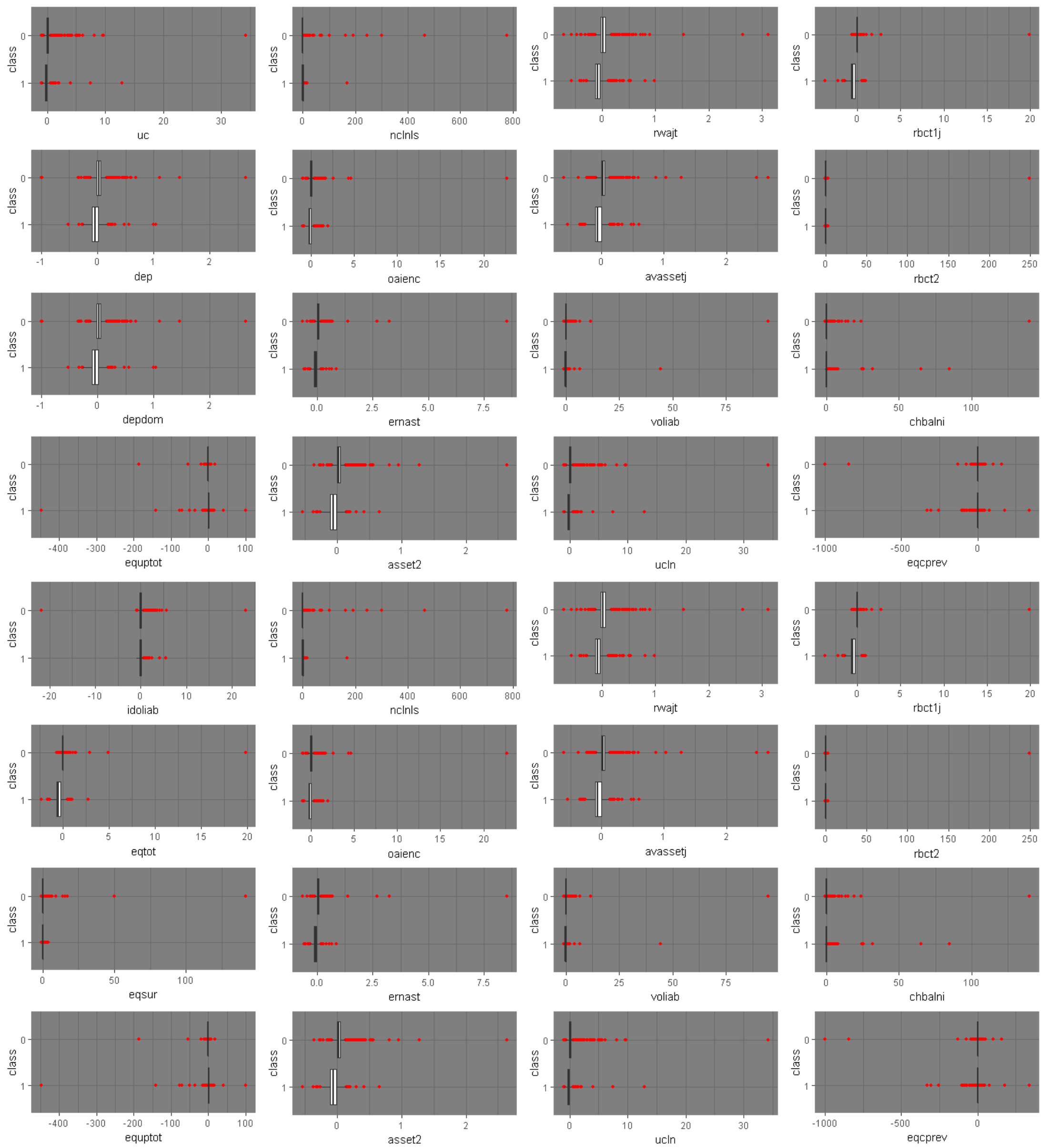


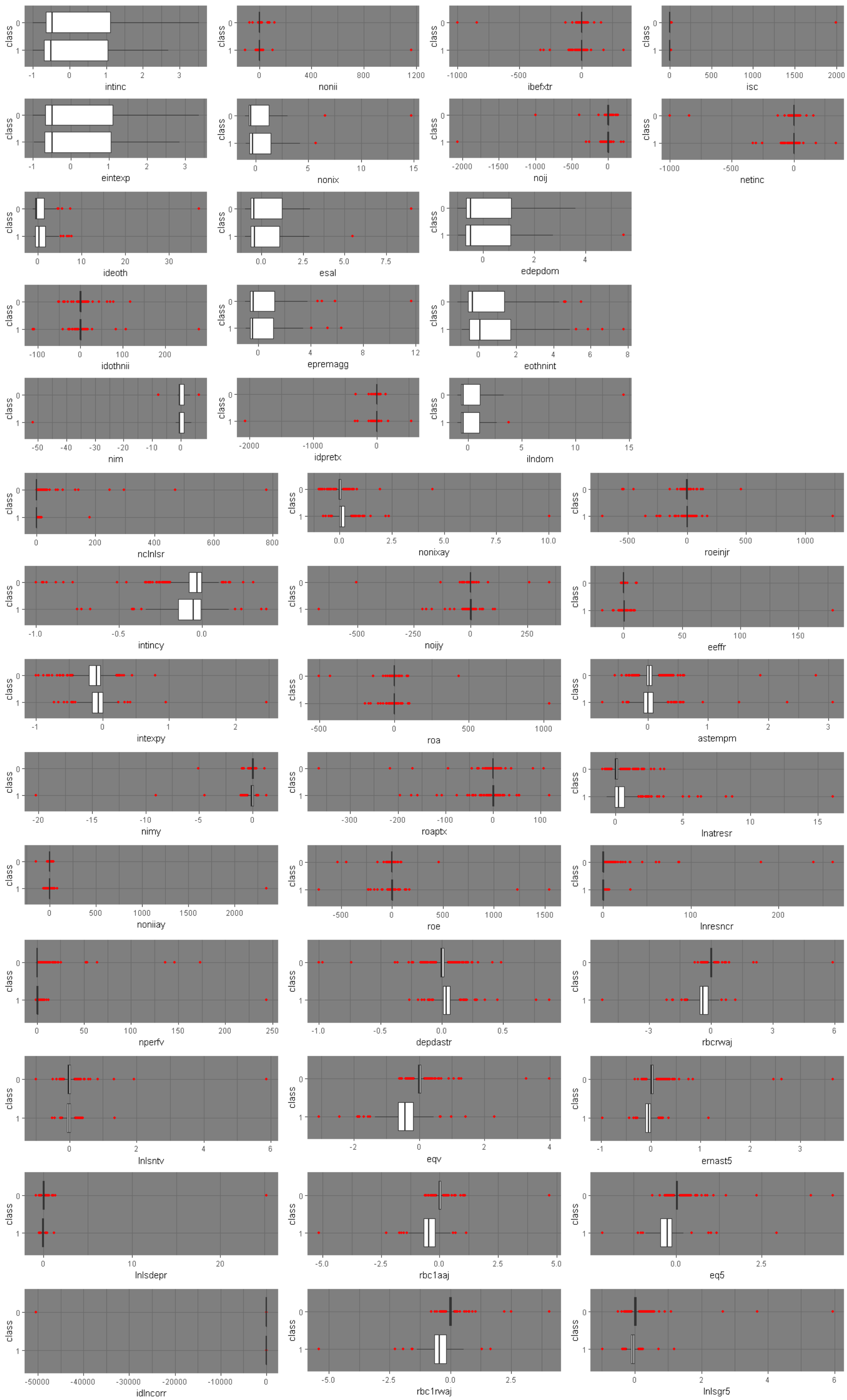

intinc
nonii
ibefxtr
isc
eintexp
nonix
noij
netinc
ideoth
esal
edepdom
idothnii
epremagg
eothnint
nim
idpretx
ilndom
nperfv
depdastr
rbcrwaj
lnlsntv
eqv
ernast5
lnlsdepr
rbc1aaj
eq5
idlncorr
rbc1rwaj
lnlsgr5
nperfv
depdastr
rbcrwaj
lnlsntv
eqv
ernast5
lnlsdepr
rbc1aaj
eq5
idlncorr
rbc1rwaj
lnlsgr5

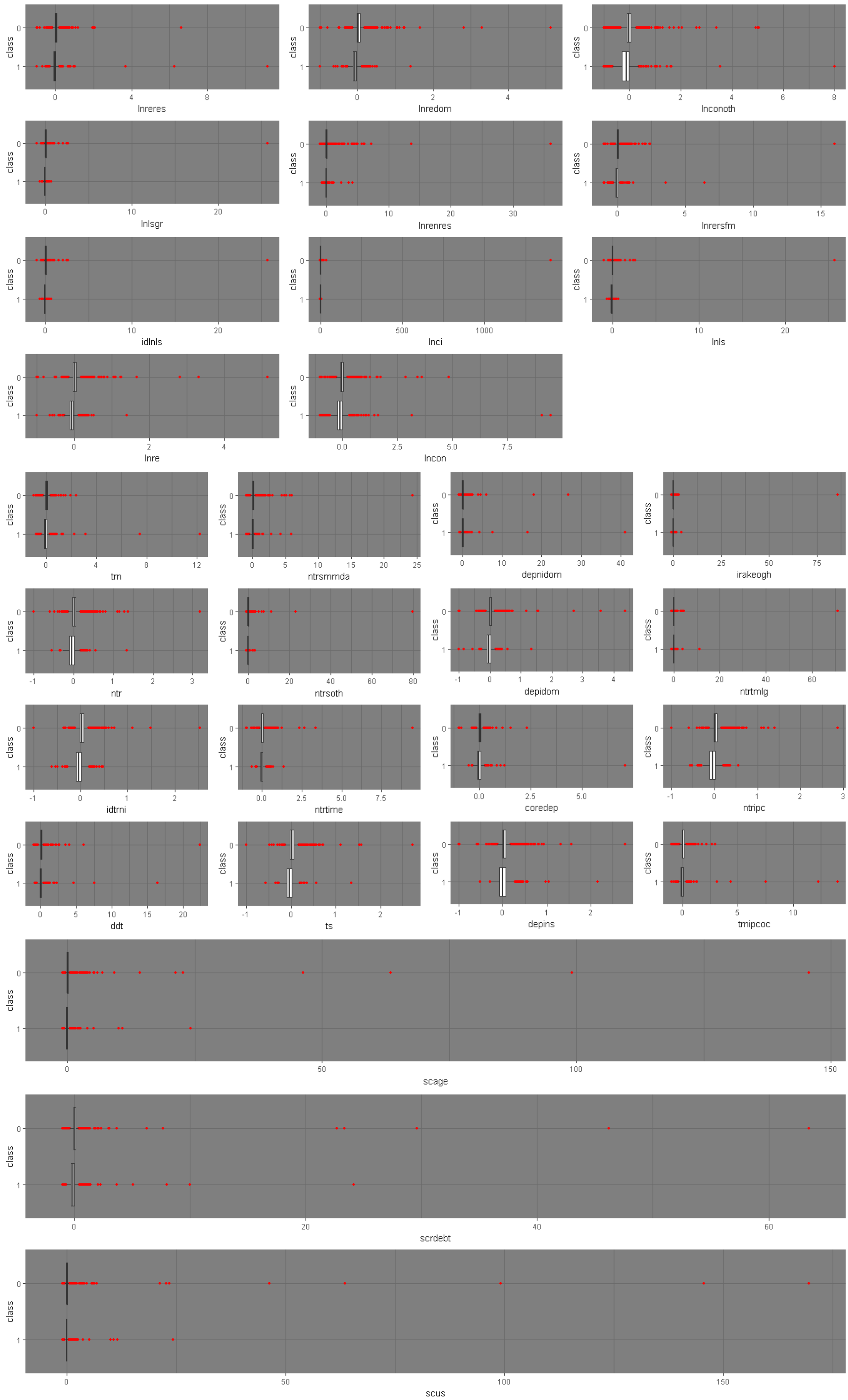

Distribution of data
class_2
1
0
lnreres
lnredom
lnconoth
lnlsgr
lnrenres
lnrersfm
idlnls
lnci
lnls
lnre
lncon
trn
ntrsmmda
depnidom
irakeogh
ntr
ntrsoth
depidom
ntrtmlg
idtrni
ntrtime
coredep
ntripc
ddt
ts
depins
trnipcoc
scage
scrdebt
scus

Appendix F

Boxplots: half-year data

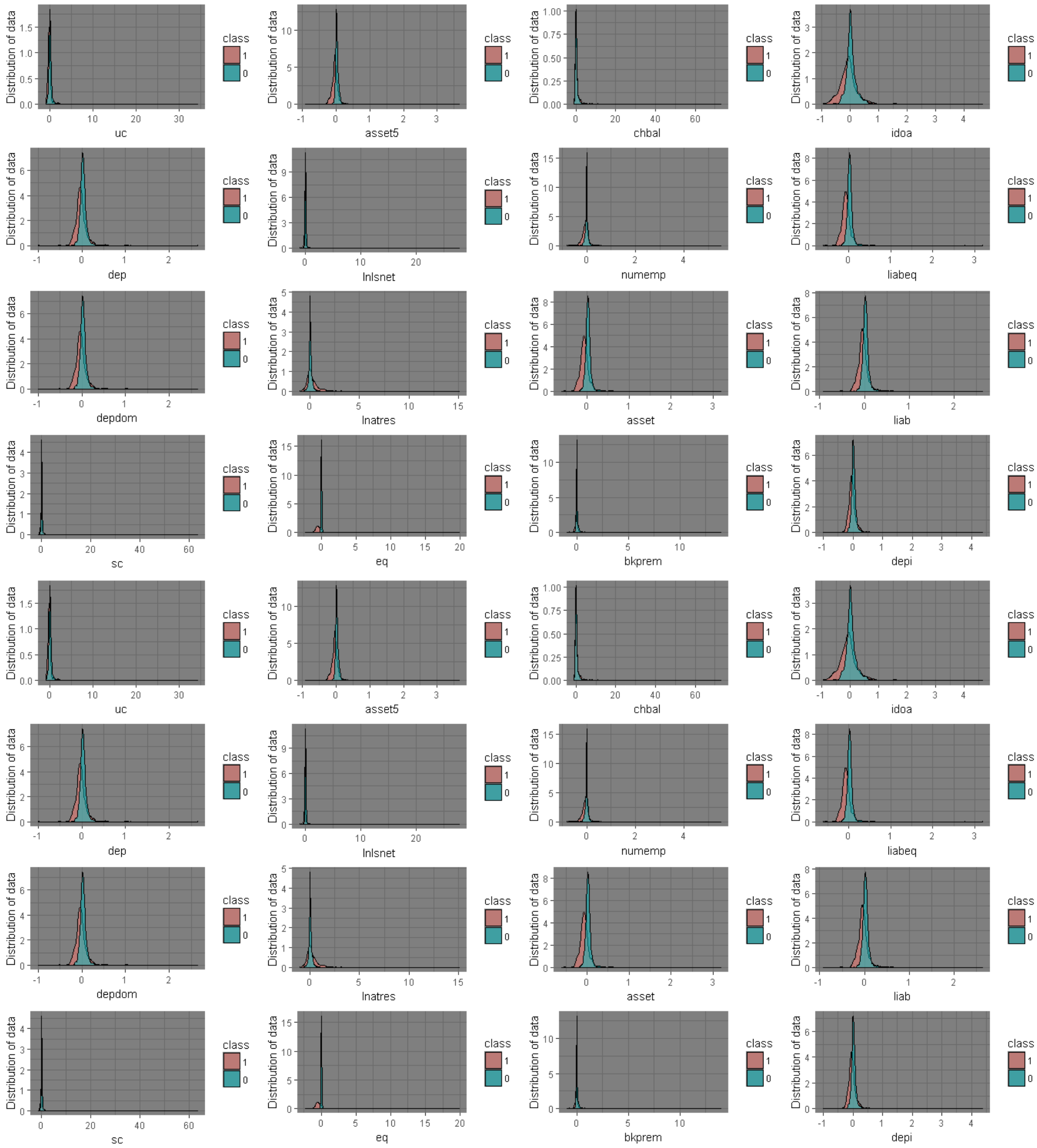


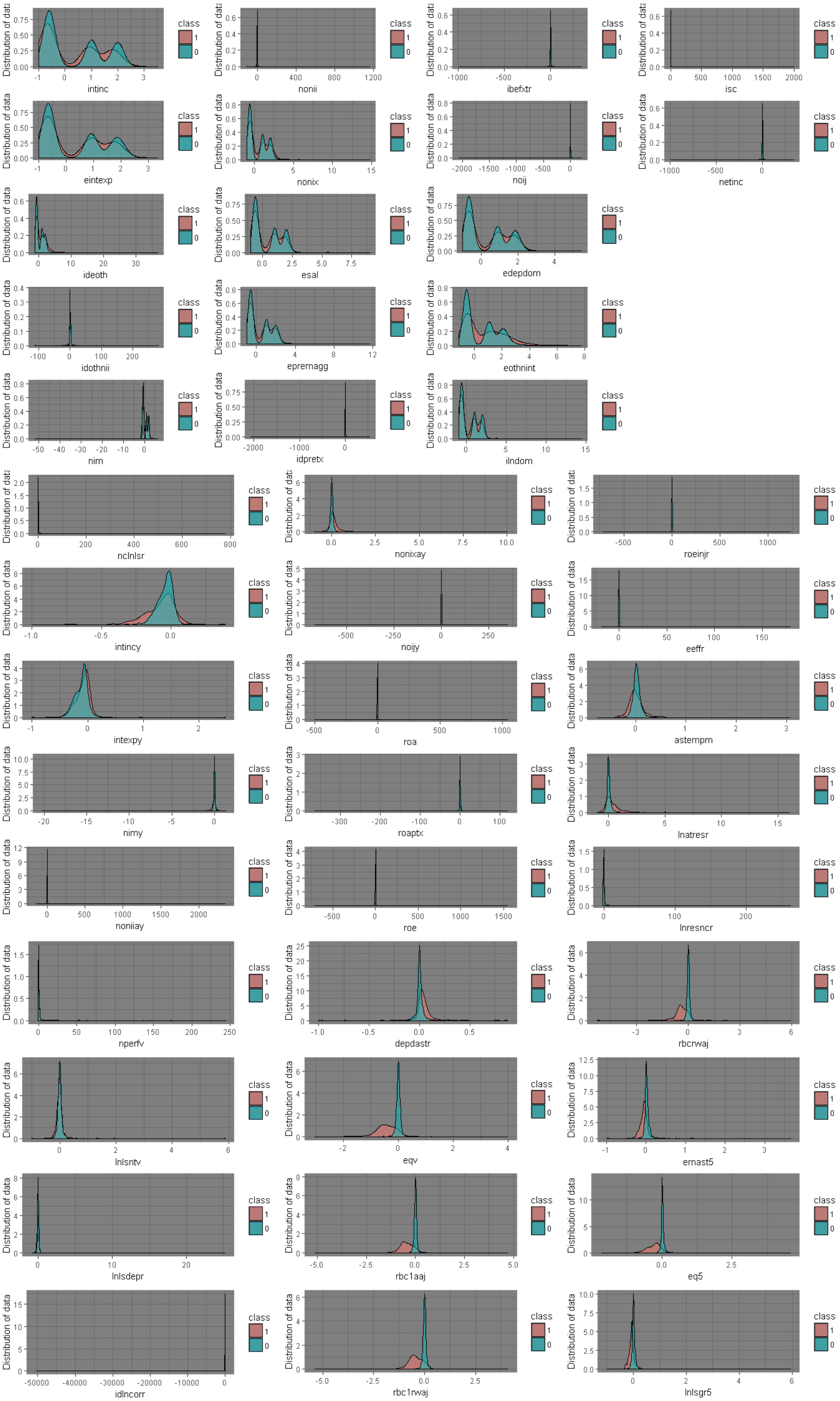

class
intinc
nonii
ibefxtr
isc
eintexp
nonix
noij
netinc
ideoth
esal
edepdom
idothnii
epremagg
eothnint
nim
idpretx
ilndom
nclnlsr
nonixay
roeinjr
intincy
noijy
eeffr
intexpy
roa
astempm
nimy
roaptx
lnatresr
noniiay
roe
lnresncr
nperfv
depdastr
rbcrwaj
lnlsntv
eqv
ernast5
lnlsdepr
rbc1aaj
eq5
idlncorr
rbc1rwaj
lnlsgr5

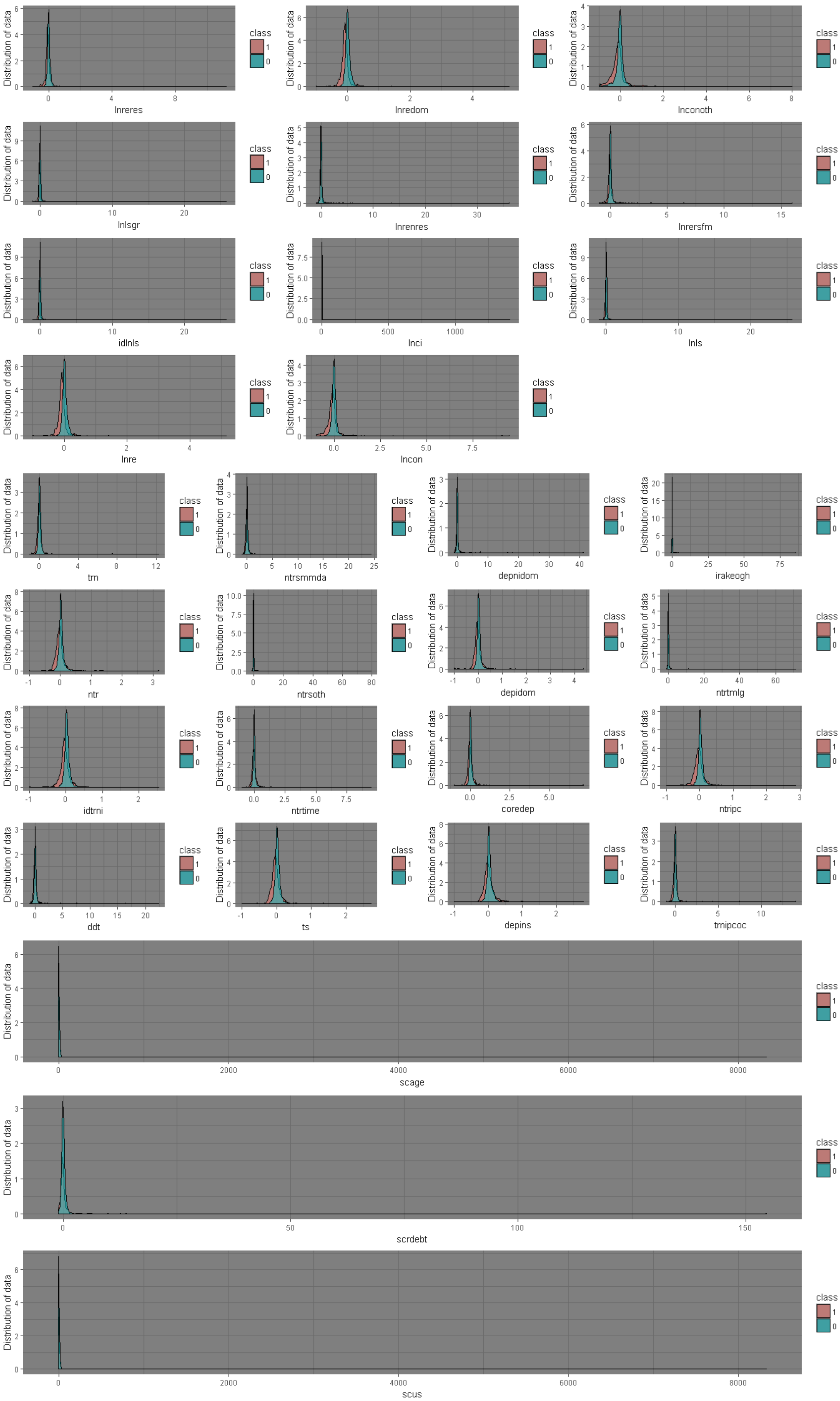

lnreres
lnredom
lnconoth
lnlsgr
lnrenres
lnrersfm
idlnls
lnci
lnls
lnre
lncon
trn
ntrsmmda
depnidom
irakeogh
ntr
ntrsoth
depidom
ntrtmlg
idtrni
ntrtime
coredep
ntripc
ddt
ts
depins
trnipcoc
scage
scrdebt
scus
class

Appendix G

Density plots: half-year data

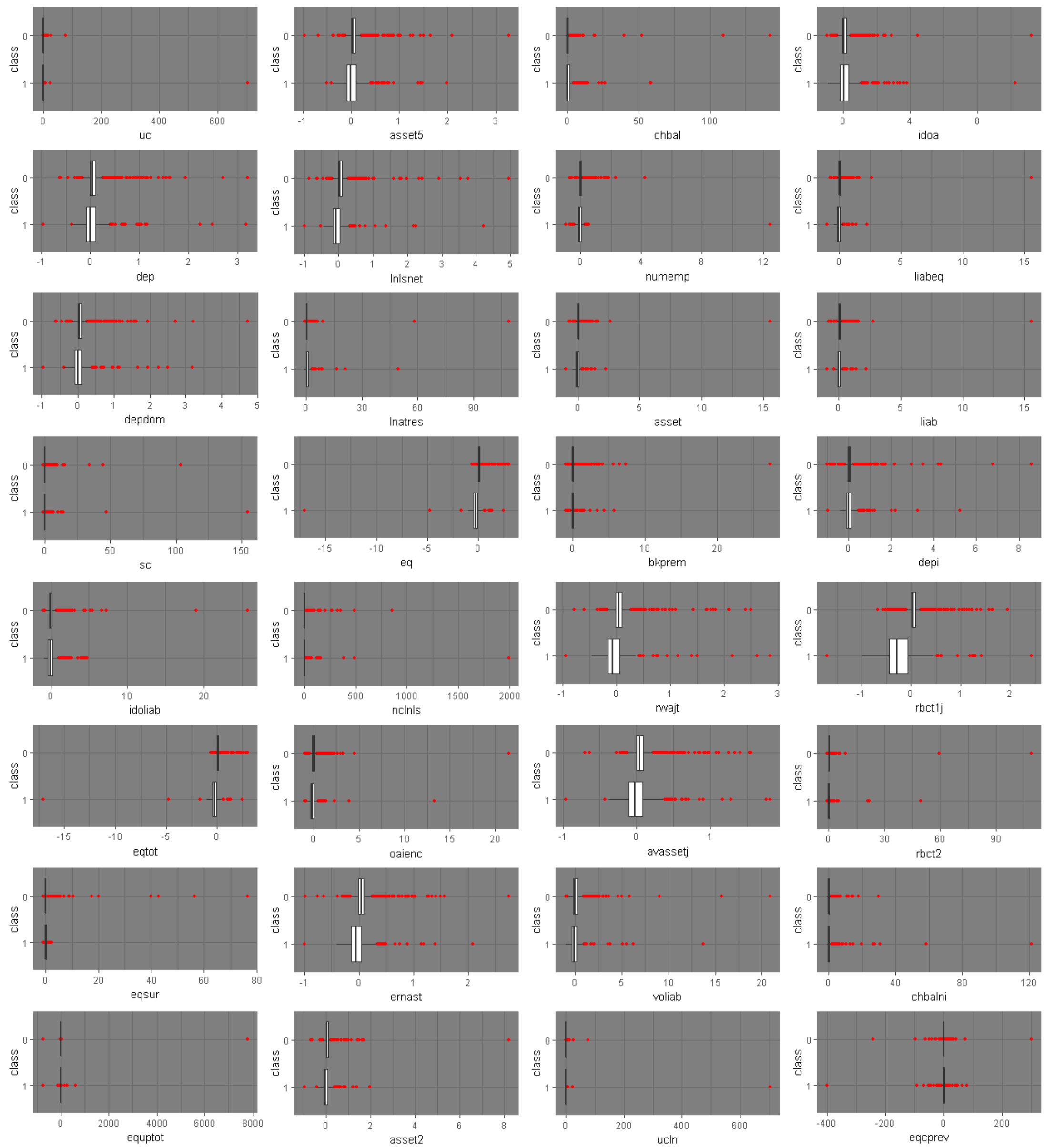


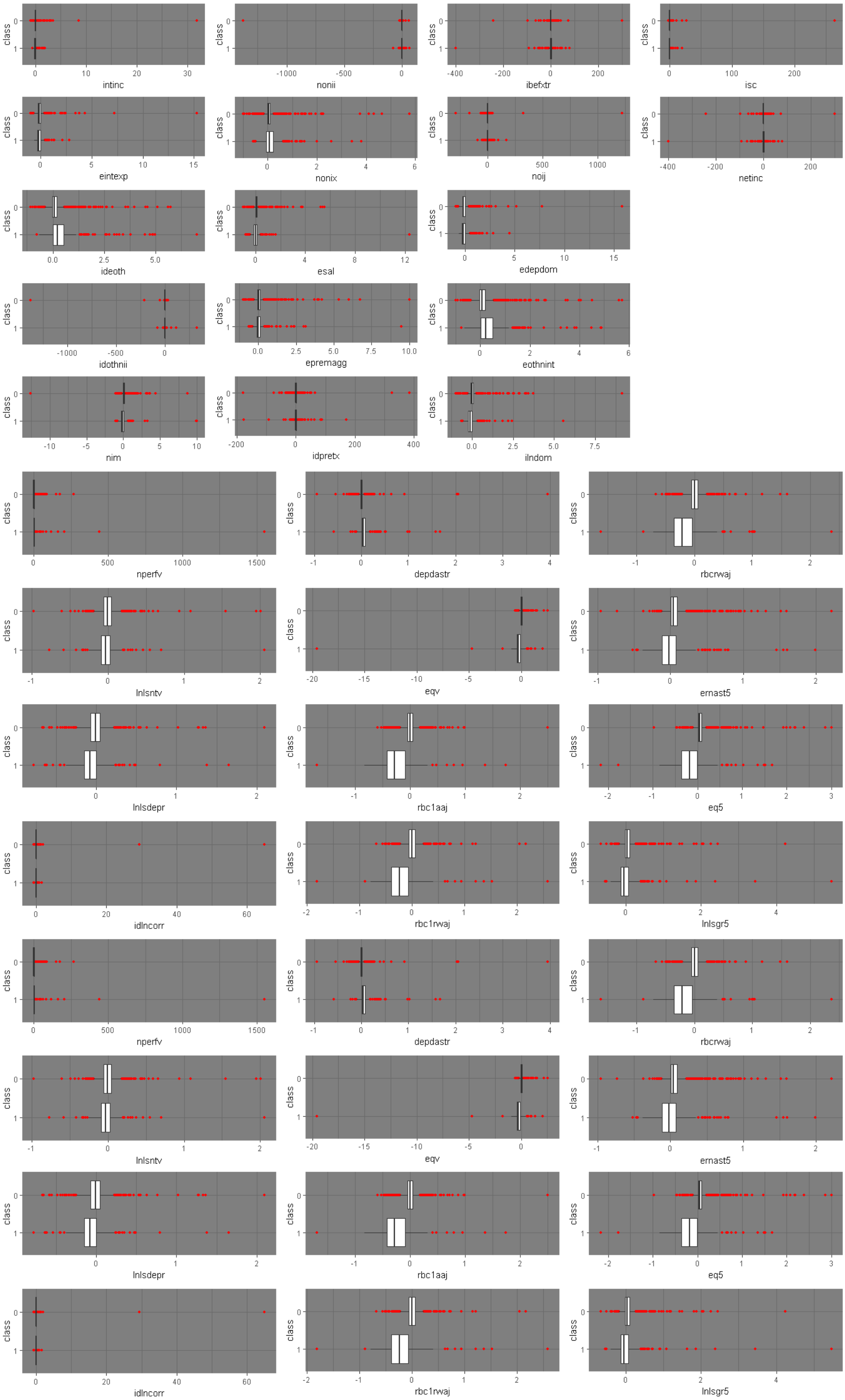

Distribution of data
class
1
0
intinc
nonii
ibefxtr
isc
eintexp
nonix
noij
netinc
ideoth
esal
edepdom
idothnii
epremagg
eothnint
nim
idpretx
ilndom
nclnlsr
nonixay
roeinjr
intincy
noijy
eeffr
intexpy
roa
astempm
nimy
roaptx
lnatresr
noniiay
roe
lnresncr
nperfv
depdastr
rbcrwaj
lnlsntv
eqv
ernast5
lnlsdepr
rbc1aaj
eq5
idlncorr
rbc1rwaj
lnlsgr5

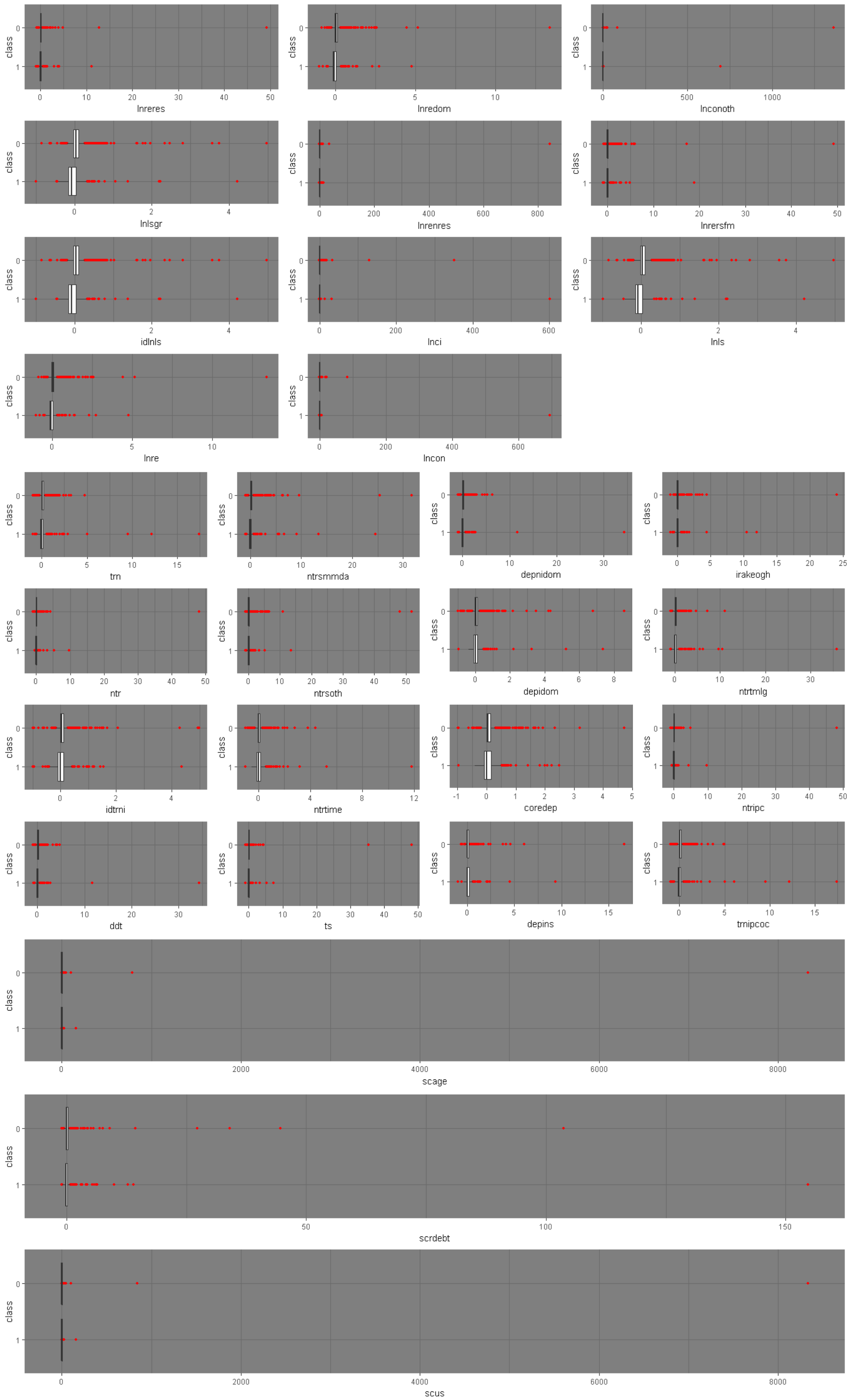

Distribution of data
class
1
0
lnreres
lnredom
lnconoth
lnlsgr
lnrenres
lnrersfm
idlnls
lnci
lnls
lnre
lncon
trn
ntrsmmda
depnidom
irakeogh
ntr
ntrsoth
depidom
ntrtmlg
idtrni
ntrtime
coredep
ntripc
ddt
ts
depins
trnipcoc
scage
scrdebt
scus

Appendix H

Boxplots: yearly data

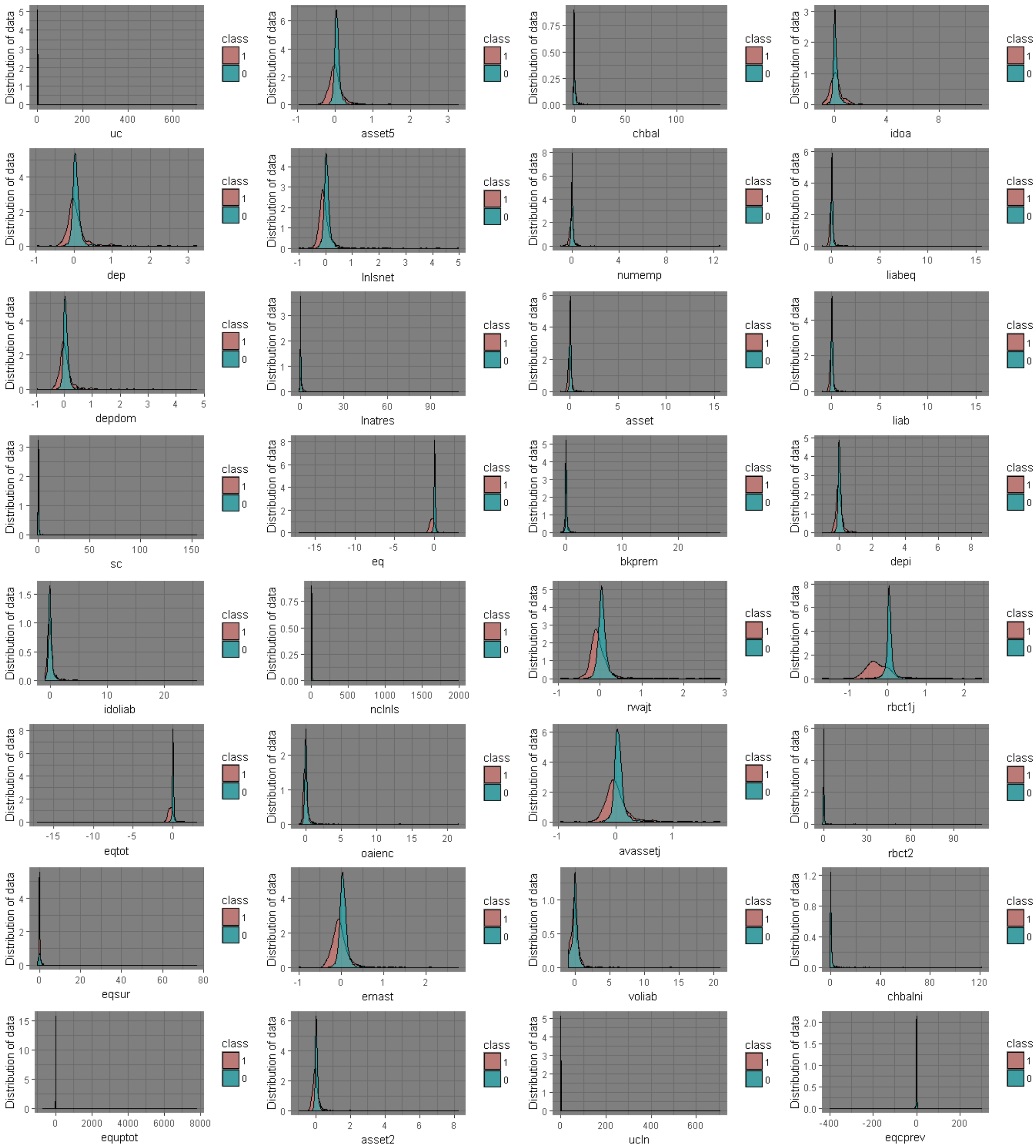


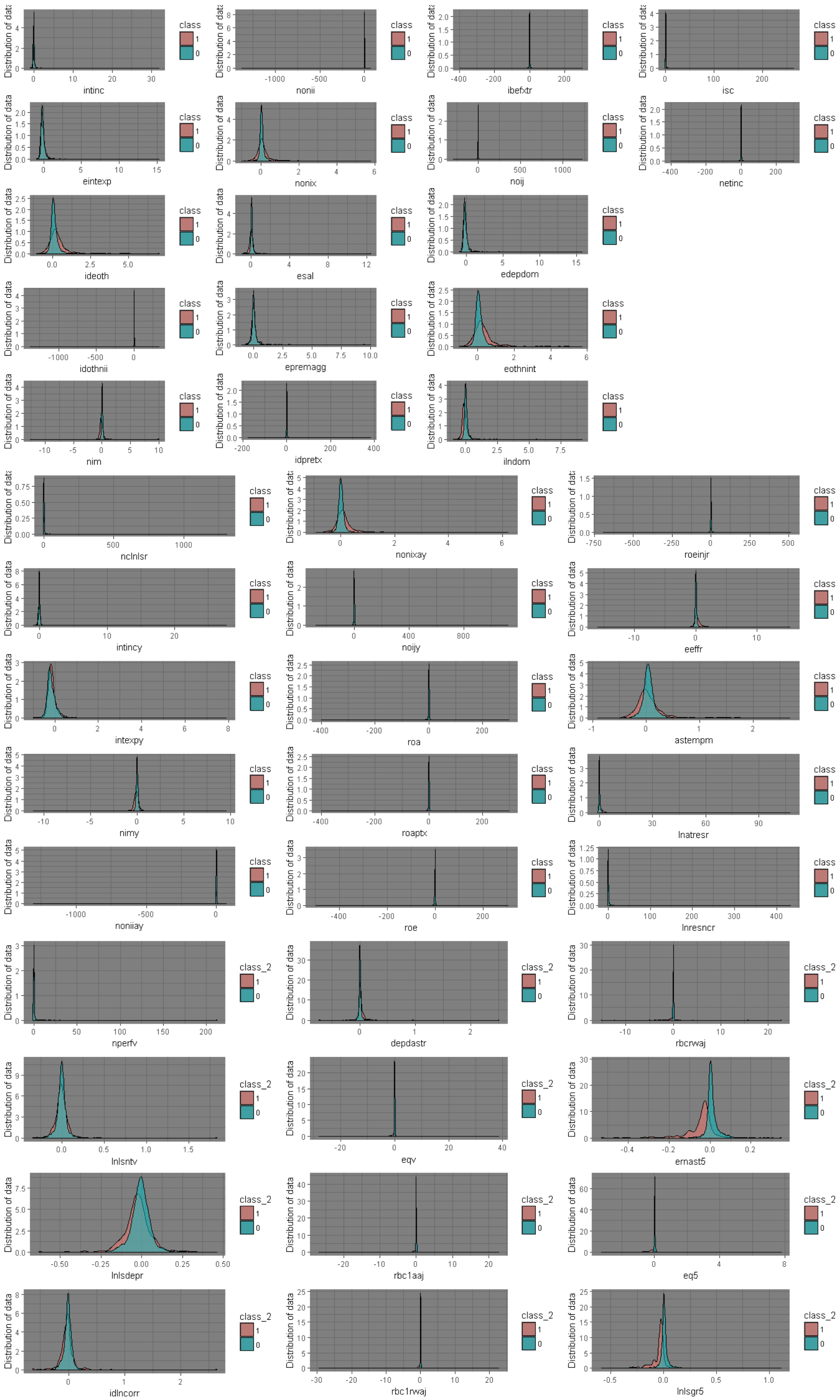

intinc
nonii
ibefxtr
isc
eintexp
nonix
noij
netinc
ideoth
esal
edepdom
idothnii
epremagg
eothnint
nim
idpretx
ilndom
nperfv
depdastr
rbcrwaj
lnlsntv
eqv
ernast5
lnlsdepr
rbc1aaj
eq5
idlncorr
rbc1rwaj
lnlsgr5
nperfv
depdastr
rbcrwaj
lnlsntv
eqv
ernast5
lnlsdepr
rbc1aaj
eq5
idlncorr
rbc1rwaj
lnlsgr5

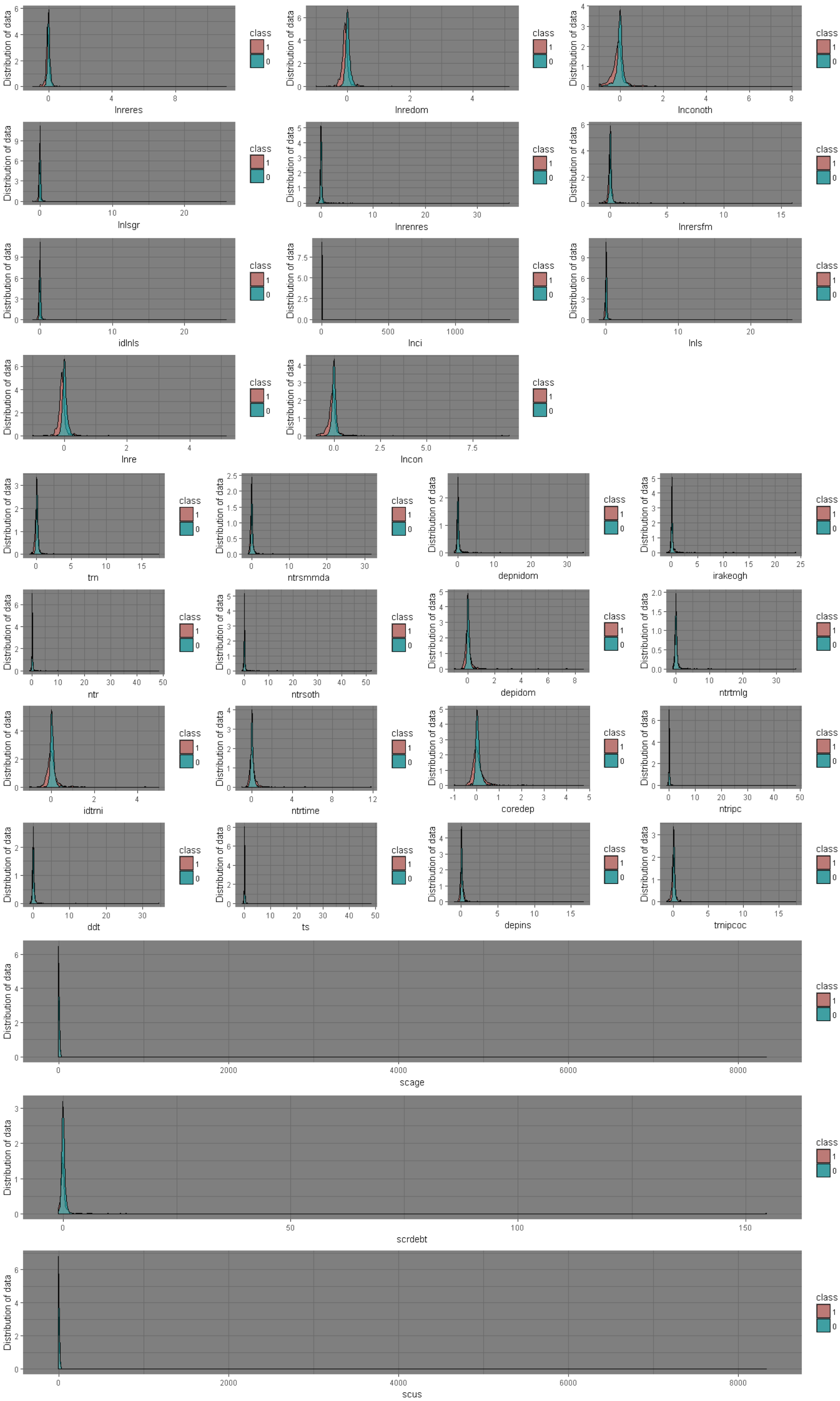

lnreres
lnredom
lnconoth
lnlsgr
lnrenres
lnrersfm
idlnls
lnci
lnls
lnre
lncon
trn
ntrsmmda
depnidom
irakeogh
ntr
ntrsoth
depidom
ntrtmlg
idtrni
ntrtime
coredep
ntripc
ddt
ts
depins
trnipcoc
scage
scrdebt
scus
class

Appendix I

Density plots: yearly data




Distribution of data
class
class_2
intinc
nonii
ibefxtr
isc
eintexp
nonix
noij
netinc
ideoth
esal
edepdom
idothnii
epremagg
eothnint
nim
idpretx
ilndom
nclnlsr
nonixay
roeinjr
intincy
noijy
eeffr
intexpy
roa
astempm
nimy
roaptx
lnatresr
noniiay
roe
lnresncr
nperfv
depdastr
rbcrwaj
lnlsntv
eqv
ernast5
lnlsdepr
rbc1aaj
eq5
idlncorr
rbc1rwaj
lnlsgr5


Distribution of data
class
1
0
lnreres
lnredom
lnconoth
lnlsgr
lnrenres
lnrersfm
idlnls
lnci
lnls
lnre
lncon
trn
ntrsmmda
depnidom
irakeogh
ntr
ntrsoth
depidom
ntrtmlg
idtrni
ntrtime
coredep
ntripc
ddt
ts
depins
trnipcoc
scage
scrdebt
scus